UNIVERSITY OF CALIFORNIA

Santa Barbara

Rethinking How Proteins Move Along DNA

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemistry

by

Adam Jacob Pollak

Committee in charge:

Professor Norbert Reich, Chair

Professor Frederick Dahlquist

Professor Stanley Parsons

Professor Song-I Han

Professor David Low

December 2014

The dissertation of Adam Jacob Pollak is approved.

_____

Frederick Dahlquist

_____

Stanley Parsons

_____

Song-I Han

_____

David Low

_____

Norbert Reich, Committee Chair

October 2014

Rethinking How Proteins Move Along DNA

Adam Jacob Pollak

ACKNOWLEDGEMENTS

I would like to thank my P.I., Norbert Reich, for providing me with invaluable guidance, knowledge, and inspiration; you are in many ways responsible for my current and future successes, and I hope we can remain colleagues and friends for years to come. I am grateful to the members of the Reich lab for your questions, answers, honesty, knowledge, jokes, political opinions, love of the outdoors, appreciation of food, willingness to exercise, need for coffee breaks, and so much more; those I would like to thank in particular include: Stephanie Coffin, Andrew Bonham, Elysia Cohn, Gary Braun, Stacey Peterson, Doug Matje, Celeste Holtz-Schietinger, Clay Woodcock, Dean Morales, Xiao Huang, Erin Morgan, Aaron Chin, Brigitte Naughton, and many others. I would like acknowledge the members of my committee, as well as our collaborators, including: Frank Brown, Itay Barel, Song-I Han, and Stacey Peterson. I have been lucky to have amazing teachers throughout my life. Obviously, none of my success and growth would be possible without my supportive family and Michelle Kem.

VITA OF ADAM JACOB POLLAK

October 2014

Education:

Ph.D., Chemistry and Biochemistry, September 2009-October 2014 (expected), University of California at Santa Barbara, Santa Barbara, CA

B.S., Chemical Biology, August 2004-May 2008, University of California at Berkeley, Berkeley, CA

Research Experience:

University of California-Santa Barbara
2009- October 2014 (expected)
Research advisor: Dr. Norbert O. Reich
Principal areas of study: enzyme kinetics, pre-steady state kinetics, enzyme mechanisms, DNA methyltransferases, restriction endonucleases, epigenetics, processivity, molecular biology, assay development.

University of California-Berkeley
2007-2008
Research advisor: Dr. Judith Klinman
Principal areas of study: enzyme kinetics, kinetic isotope effects.

Publications:

Pollak, A.J., Reich, N.O. (2012) Proximal recognition sites facilitate intrasite hopping by DNA adenine methyltransferase: a mechanistic exploration of epigenetic gene regulation. *J. Biol. Chem.* **27**, 22873-81.

Pollak, A.J., Chin, A.T.*, Brown, F. L., Reich, N.O. (2014) DNA looping provides for "intersegmental hopping" by proteins: a mechanism for long-range site localization. *J. Mol. Biol.* DOI: 10.1016/j.jmb.2014.08.002

Pollak, A.J., Chin, A.T.*, Reich, N.O. (2014) Distinct and complementary search mechanisms by *E. coli* Type II restriction endonucleases. *Biochemistry (accepted)*

Pollak, A.J., Reich, N.O. (2014) The role of preexisting protein-DNA "roadblock" complexes in the search mechanisms by sequence-specific DNA binding proteins. *J. Mol. Biol. (under review)*

*mentored undergrad

Conferences:

Enzyme Mechanisms Conference Coronado, CA: Poster presentation (January 2013)
Gordon Research Conference: Enzymes, Coenzymes, and Metabolic Pathways Waterville Valley, NH: Poster presentation (July 2014)

Teaching experience:
UCSB
Chem1A, B, C; Chem112L (2009-2013)
Mentored ~12 different undergrads/rotation students in the Reich Lab

Press releases:
Santa Barbara News Press
"UCSB researchers work to eliminate bacterial infection." (July 29, 2012)

UCSB Press Release
"UCSB Researchers' Discovery of 'Intrasite Hopping' of Bacterial Enzyme on DNA Gives Insight into Epigenetic Gene Expression." (July 17, 2012)

Outreach:
UCSB
Sci-Trek: taught science modules to at-risk 7th graders. (2013-2014)

Work Experience
Company: Corium International, Menlo Park, CA (2008-2009)
Title: Research Associate
Duties: Developed and utilized HPLC based methods to detect and quantify protein drug degradation patterns.

References:
Norbert Reich (UCSB Professor, Dept. of Chemistry and Biochemistry):
reich@chem.ucsb.edu, (805) 893-8368

Frederick Dahlquist (UCSB Professor and Chair, Dept. of Chemistry and Biochemistry):
dahlquist@chem.ucsb.edu, (805) 893-5326

ABSTRACT

Rethinking How Proteins Move Along DNA

by

Adam Jacob Pollak

Site-specific DNA binding by proteins is critical for diverse biological processes, including gene regulation, DNA repair, DNA replication, immune response, and others. In most cases, this binding is preceded by facilitated diffusion, where the protein passively moves along nonspecific DNA (DNA lacking particular binding sites) in search of its site(s). Details concerning the mechanisms of how proteins move along DNA remain unclear yet actively investigated. Several reports claim that facilitated diffusion relies primarily on sliding and hopping processes, where the protein moves along the trajectory of the DNA helix. However, these mechanisms involve redundant searches over local regions (up to hundreds of base pairs), while site finding *in vivo* requires searches of hundreds of thousands of base pairs. We "rethink" facilitated diffusion in two ways: 1) what are the mechanisms available for facilitated diffusion? And 2) how are these mechanisms driven by proteins' specific biological context?

We investigate the facilitated diffusion properties of *E. coli* DNA adenine methyltransferase (Dam), which methylates palindromic 5'-GATC-3' sites on the N^6 position of adenine. We use an *in vitro* steady-state processivity assay, where the ability of the enzyme to catalyze multiple methylations within a single binding event is quantified. Our results are suggestive of a new mechanism, intersegmental hopping, where proteins hop

between regions of a single DNA molecule that are looped together. This mechanism allows for efficient long-range movements, and is similar to other ambiguously described looping mechanisms proposed in recent reports. Dam's low cellular copy number and need to move along the entire bacterial chromosome (i.e. its biological context) is ideally accommodated by this mechanism, as we imagine will be true for other proteins with low copy numbers and rare sites.

We also demonstrate that EcoRI ENase uses an extreme sliding mechanism, previously hypothesized to be unlikely. EcoRI ENase and EcoRV ENase both cut incoming phage DNA as part of type II restriction modification systems, and are shown here to use distinct facilitated diffusion mechanisms. We argue that their site finding mechanisms are complimentary, ensuring robust phage defense, again demonstrating the connection between facilitated diffusion and biological context.

Furthermore, we show that Dam is able to move along DNA in spite of pre-incubated protein roadblocks. Provocatively, DNA-loop inducing roadblocks, such as the histone-like Lrp, can improve the translocation process by specifically enhancing intersegmental hopping. Therefore, this mechanism is proposed as a general way for proteins to move along the crowded, compacted genomic DNA landscape.

Dam was previously demonstrated to undergo intrasite processivity, where the methylation of both strands within a single GATC site proceeds by a single Dam enzyme dramatically reorienting itself while switching between strands. Here, we show that intrasite processivity is only possible on long stretches of DNA when GATC sites are clustered. GATC sites are clustered in most known and putative Dam methylation dependent epigenetic operons, which involve transitions from uniquely unmethylated GATC sites to

fully methylated ones. Intrasite processivity is therefore postulated as necessary for epigenetic gene regulation by Dam.

Table of Contents:

List of Figures:

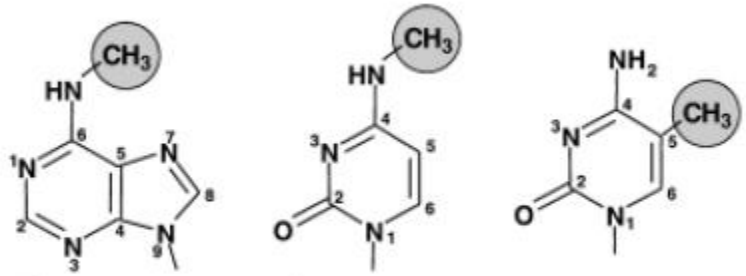
List of Tables:

I. Introduction

1.1 DNA methylation

In spite of its minimal manipulations to DNA's overall structure (1), the methylation tag on DNA provides a variety of essential information for the cell. The intricacies and complexities of the consequences of these methylation signals are diverse and widespread throughout all domains of life. DNA methylation expands the scope of genetic information from the four canonical bases: adenine, cytosine, thymine, and guanidine. DNA methylation of cytosine (C^5-methyl or N^4-methyl) or adenine (N^6-methyl) within particular sequences encompasses the breath of biologically occurring methylated DNA bases (although in some cases these can be further functionalized) (1,2) (Figure 1.1). The appreciated biological and biomedical significance of DNA methylation is rapidly expanding, and understanding fundamental aspects of the methyltransferases and their tightly regulated functions is critical (3).

Figure 1.1. Enzymatically modified bases in DNA. This figure was used with permission from John Wiley and Sons (2).

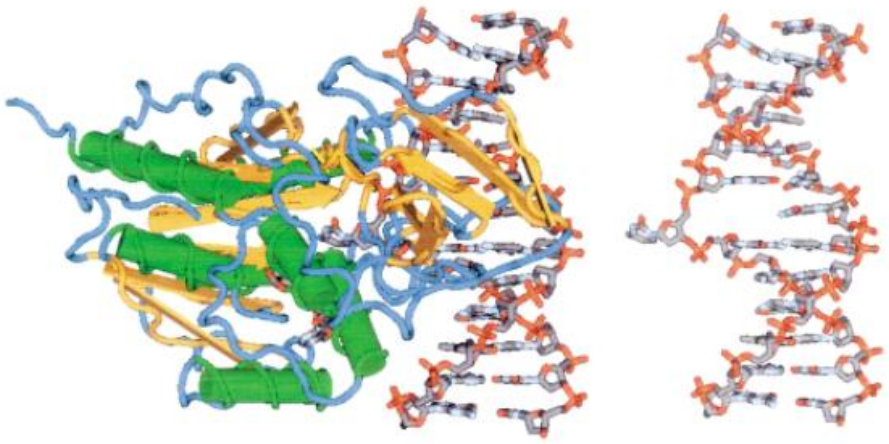


N^6-methyladenine N^4-methylcytosine 5-methylcytosine

DNA methylation was originally discovered in 1925 (4), and further mentioned in the literature in 1948 (5). Its involvement in higher order cellular function was hypothesized in the 1960's (1). The first DNA methyltransferase was discovered in 1964 in *E. coli* (6). A

defining feature of DNA methyltransferases (and most other enzymes that modify DNA, such as in DNA repair enzymes) is that they flip their target base out of the DNA duplex prior to the methyl transfer event (7) (Figure 1.2). This energetically costly removal of the base stacking interactions is usually mitigated by the protein inserting a loop into the disordered helix (Figure 1.2). Interestingly, the rearrangement of the helix upon base flipping in many cases provides an indirect read-out source of specificity (2). Physically, the methylation of DNA acts to either block the binding of DNA regulatory proteins, or to recruit certain proteins that specifically bind to methylated DNA (8). Also, methylated DNA bases manipulate the curvature of DNA and lower its thermodynamic stability (9), which has been shown to have additional consequences, particularly in the initiation of bacterial replication (1). Methyltransferases utilize the universal methyl donor SAM (S-adenosylmethionine) as the source of the methyl moiety (10). The SAM binding pocket is conserved in all methyltransferases, including those that methylate proteins, and RNAs (1).

Figure 1.2. The structure of the M.HhaI methyltransferase-DNA complex (left). The DNA structure with the base flipped out without the protein bound (right). This figure was used with permission from John Wiley and Sons (2).



In humans, 5-methylcytosine methylation occurs, and adenine methylation is not found (Figure 1.1). Methylation is quite extensive in the human genome in that 60-90 % of palindromic 5'-CG-3' sites are methylated (2,11). *De novo* methylation (the initial methylation of a completely unmethylated site), maintenance methylation (methylation of a hemimehtylated site), and demethylation (the removal of methylation marks) pathways together create an elaborate and tightly regulated system that governs the methylation state of the human DNA (11). The regulation and dramatic fluctuation of the *de novo* methyltransferase expression is most consequential. DNA methylation has many gene regulatory functions. For example, it mediates genomic imprinting and x-chromosome inactivation (11). Many stages of development, starting with the initial fertilization event involve considerable changes in DNA *de novo* methyltransferase levels (12). As a consequence, methylation patterns are remarkably different between different types of tissues (13). Importantly, the maintenance methyltransferases cannot change the methylation patterns. Aberrant DNA methylation patterns in many forms of cancer are the basis for a growing focus on the epigenetic contributions to diseases (14). Direct targeting of epigenetic enzymes, such as small molecule DNA methyltransferase inhibitors and histone deacetylase inhibitors are fast becoming a part of the pharmaceutical industry's therapeutic arsenal. Some DNA methyltransferase inhibitors, such as 5-azacytidine, work by being incorporated into the genomic DNA, then act as suicide inhibitor to the methyltransferases.

1.2 Epigenetics

Epigenetics can be loosely defined as molecules independent of the 4 canonical bases (above) that are heritable and have a role in the expression (or lack thereof) of genes (15); such molecules can include modified DNA, siRNA, histones, and more The simplest

molecular example involves methylated DNA. Due to all living cells' reliance on semi-conservative replication, modifications made to the parental strand will persist in the daughter cell, and this information (which is independent of the 4 canonical bases) in many cases is involved in the regulation of genes, and serves as a template for methylation of the daughter strand, allowing maintenance of the signal through subsequent cell divisions (16). An example of epigenetics on the level of phenotype involves the Dutch famine of 1944, where millions of people were nearly starved (17). Interestingly, the grandchildren of the survived starved individuals were smaller than average, which suggested that the famine caused changes in the epigenetic makeup that was inherited though multiple generations.

Touted as a new approach to understanding molecular biology, particularly in the context of its relation to human disease, epigenetics has recently received an explosion of interest (15). However, prior to a sufficient appreciation of the work of Mendel in the 1860's, what would be described as epigenetics today was certainly significantly considered in the place of what became classical Mendelian genetics (18). The work of Mendel, which is the basis of the field of genetics, assumes that environmental changes have no effect on the expression of genes of newly propagated cells, an idea that only became popular roughly 50 years after Mendel's first work on genetics (19). The notion of, for example, what you eat and what toxins you expose your body to will have an effect on the gene expression of propagated cells and progeny, is popular today because of a "new" appreciation of epigenetics, and was interestingly thought to be important prior to an appreciation of Mendel's work.

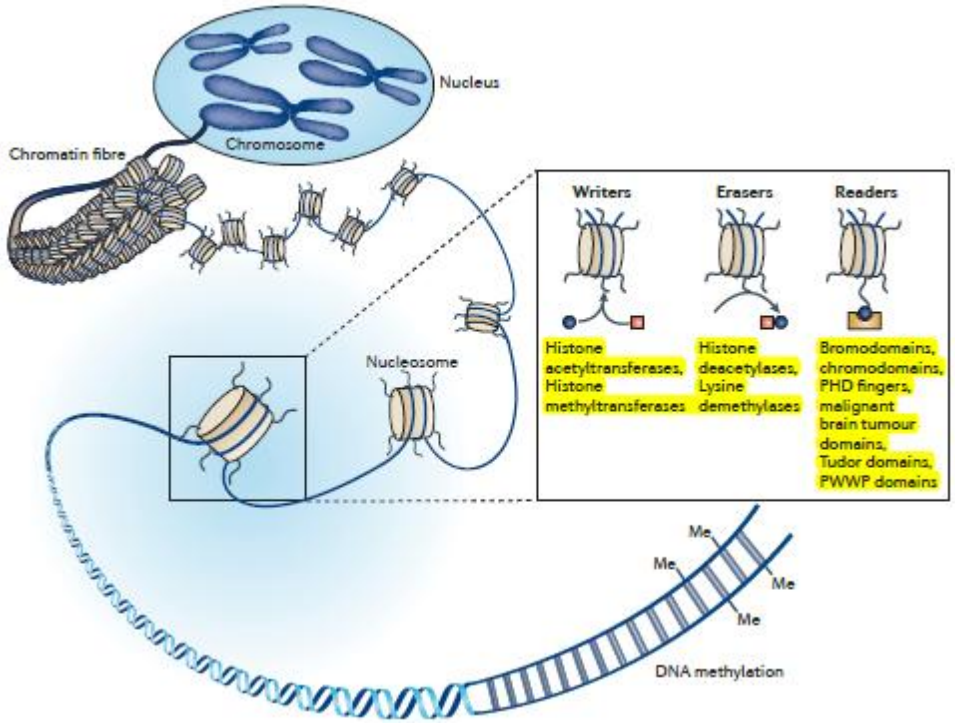
Following the discovery of the DNA double helix in the 1950's by Watson, Crick, and Franklin, an understanding and appreciation of genetics was rapidly growing, forming the

basis of our current understanding of molecular biology. However, the work of developmental biologists in particular revealed inconsistencies in the central dogma of molecular biology. For example, the ability of genetically identical cells to remain differentiated in *Drosophila*, i.e. the maintenance of eye cells to eye cells and wing cells to wing cells, remained unexplained by classical genetics (18). In an attempt to reconcile such inconsistencies, Conrad Waddington and others proposed that there must be a connection between the idea of genetics and developmental biology, and combined the term "epigenisis," which concerns cell differentiation, with genetics, and coined "epigenetics" in 1939 (18,20). While this definition continues to be altered with better molecular understandings of the relevant processes, the original definition of epigenetics by Waddington: "the unfolding of the genetic program for development" remains valid (20,21).

The epigenetic role of DNA methylation-dependent gene silencing in mammals acts in concert with other epigenetic phenomenon (22). The massively large human genome is heavily compacted by eukaryotic proteins called histones (23). Particularly tight compaction by histones results in the inability of the transcriptional machinery to transcribe DNA, and is associated with high levels of DNA methylation (24). This is referred to as heterochomatin, and is distinct from euchromatin, where the DNA is available to be transcribed since the wrapping of the DNA is looser around the histones (Figure 1.3). In addition to the observed methylation patterns associated with each state, other epigenetic markers certainly play a role. For example, histones are highly post-translationally modified, and particular modifications are associated with each chromatin state (Figure 1.3). There are a variety of enzymes that modify a several types of histones in many different locations on the histone protein. These include: acetyltransferases and methyltransferases, along with enzymes that

remove those groups (22) (Figure 1.3). There are also a variety of proteins that associate with different histones differentially based on their modification state. Furthermore, the production of particular ncRNAs is specific to different chromatin states (24). Both histone modifications and ncRNAs are heritably transferred between cells, in line with the definition of epigenetics. Interestingly, the details concerning the relationship and causality of these varied epigenetic components remain, in general unknown, but of outstanding interest (22-24).

Figure 1.3. DNA is wrapped around histones to form nucleosomes. DNA can be methylated, and a variety of modifications can be made to histones. These epigenetic tags regulate the transcription of genes, and are heritable. This figure was used with permission by the Nature Publishing Group.



1.3 Orphan methyltransferases

DNA methyltransferases were first discovered in the context of restriction modification systems (Chapter 3), which was initially described by work in the 1960's (25). With type II

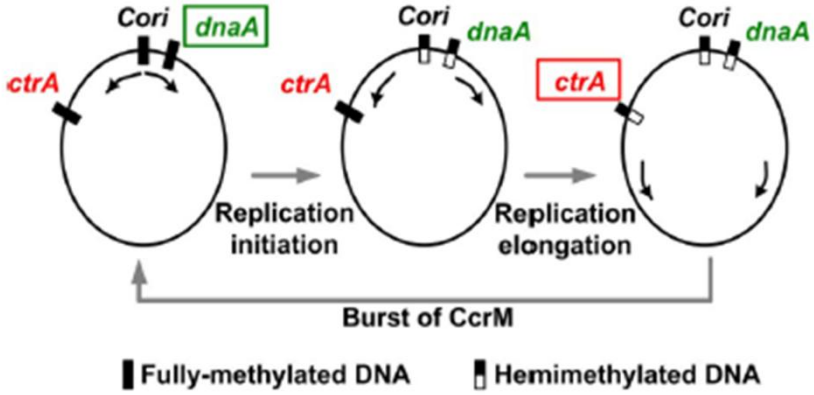
restriction modification systems, the methyltransferase has a separate restriction endonuclease partner, which recognizes and cuts DNA at the same site (26). Importantly, the endonucleases cannot cut the DNA if it is methylated by the methyltransferase. This non-gene regulatory phenomenon is used as a way for cells to protect themselves from invading nucleic acids, which are usually viruses (26), and is considered a form of a bacterial immune system. However, the biology of orphan methyltransferases (ones without a restriction endonucleases partner) is in many ways distinct and has been shown to have widespread cellular consequences distinct from the roles in restriction-modification.

The two best studied orphan methyltransferases are *E. coli* Dam and *Caulobacter crescentus* CcrM (cell cycle regulated methyltransferase) (16). Dam is involved in a variety of cellular activities and pathways (27), and we imagine this is the source of many of the unique properties of Dam discussed in this thesis. Dam is involved in mismatch repair, where immediately following replication the hemimethylated site (where the methyl group is always on the parental strand) acts as a signal for the mismatch repair machinery to edit errors in replication. Evidence for Dam's involvement in this pathway relied in part on experiments with both under (i) and over (ii) producing *E. coli* Dam strains; both strains showed an increase in mismatches as the presence of hemimethylated Dam sites decreased (27). In (i), there were no methylation marks for the mismatch repair machinery to identify as a signal. In (ii), due to the high levels of Dam, each Dam site was doubly methylated prior to the ability of the mismatch repair machinery to locate the hemimethylated sites. Dam's participation in replication is more subtle. There is a high density of Dam sites in the *oriC*, the origin of replication, and within the *DnaA* promoter (16,28). These fully-methylated Dam sites cause an increase in DnaA expression, which is required to initiate replication. In

E coli, the methylation state of these sites is used by the cell to distinguish between old and new origins. In other bacteria, the hemi-methylated sites act as a substrate for SeqA, whose binding limits re-initiation of replication by blocking DnaA binding. Dam's involvement in epigenetic gene regulation has also been well-studied (Chapter 5).

Unlike Dam, CcrM, which methylates adenines in GANTC sites, is required for *Caulobacter* viability and is expressed at only specific times in the cell cycle (28). *Caulobacter* has two cell types, stalked and swarmer (28). Replication only takes place in the swarmer cells, and the differential methylation state of several gene regulatory GANTC sites is critical for controlling replication. Interestingly, the transition from the fully to hemimethylated sites as caused by replication is essential for this process (28,29). The location of these regulatory GANTC sites on the chromosome is also important in relation to the timing and trajectory of the replication fork (29).

Figure 1.4. The methylation states of select regulatory 5'-GANTC-3' sites control the timing of DNA replication in the *Caulobacter* cell. This figure was used with permission from PNAS (29).



When the GANTC sites in the *dnaA* gene (which is close to the origin) are fully methylated, replication is initiated by DnaA; but, once replication starts, the *dnaA* gene

becomes hemimethylated, and DnaA expression is turned off (29) (Figure 1.4). The *ctrA* gene, which turns off replication, is activated when the GANTC in the promoter is hemimethylated as a consequence of replication. Importantly, the *ctrA* gene is much farther along the path of the replication fork than the *dnaA* gene (29). Finally, when the *ccrM* gene becomes hemimethylated, CcrM expression is turned on, which re-initiates the replication cycle (Figure 1.4).

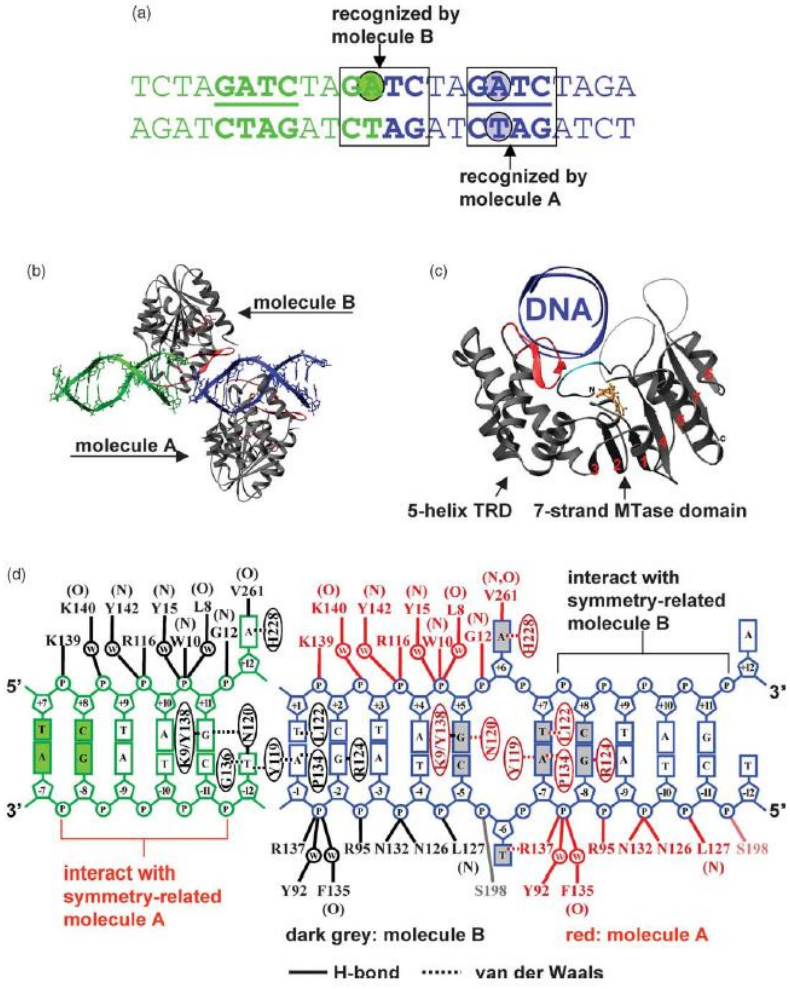
Both Dam and Ccrm represent attractive targets for anti-bacterial treatments. The lack of N^6-methyladenine in humans presents the possibility that a targeted, specific N^6-adenine methyltransferase inhibitor will have limited toxicity for humans, particularly since the methylation mechanisms between cytosine and adenine methylation are quite distinct (2). CcrM is required for *Caulobacter* survival. Dam is not required for *E. coli* survival, but has interestingly been demonstrated to be necessary for virulence in not only *E. coli*, but several other bacteria which contain Dam (16).

1.4 Dam Structure

A crystal structure of *E. coli* (Eco) Dam has been solved (30), along with the related T4 bacteriophage Dam (31). EcoDam contacts a total of 10 bp, including the 4 bp of the cognate site and the DNA directly surrounding the sites on both sides (Figure 1.5). Dam contains two domains: a 7 stranded catalytic domain with a SAM binding site, and DNA binding domain with a 5 helix bundle and a β-hairpin loop, similarly to T4 dam (30) (Figure 1.5). The DNA binding loop contains 4 conserved residues that contact the phosphates on the DNA surrounding the recognition site (R95, N126, N132, and R137) (30, 32). The use of the 2-aminopurine analog has been invaluable for understanding the details of the base flipping (33). This is due to the dramatic increase in fluorescence when the base is flipped

out of the DNA helix, which provides stabilizing base stacking (33). This technique showed that in the presence of SAM, the flipped out adenine is then stabilized by interactions with aromatic residues (which in turn decreases the flouresence) (34). Interestingly, the orphan thymine is also flipped out of the DNA helix, and this is in contrast with the T4 structure. The loss of stability of the DNA helix due to the flipped out bases is in part mitigated by the intercalation of the Y119 aromatic ring. Interestingly, the Y119 residue is necessary for base-flipping.

Figure 1.5. The structure of Dam-DNA complex A) Location of bound GATC sites in the crystal structure. B) Schematic of bound Dam proteins in the crystal structure. C) Schematic of Dam protein domains. The recognition loop is in red. D) Contacts between Dam and DNA. This figure was used with permission from Elsevier (30).



T4 dam structures have been solved for complexes with both specific and nonspecific DNA binding modes (31). These results provide a structural snapshot of how the facilitated diffusion process might occur, as this process involves extensive interactions between the protein and the nonspecific DNA. The general structure of the protein is very similar in both the specific and nonspecific binding modes (31). Interestingly, certain residues involved in specific binding are also able to bind nonspecific DNA, albeit in a completely electrostatic mode with the phosphate or sugar backbone. This observation is similar to one made for the structure of the lac repressor bound nonspecifically to DNA (35). Another notable feature of the T4 Dam structure is that the DNA binding loop is rotated nearly 80° in the nonspecific mode (31), and this is consistent with the observation that certain residues switch between specific, nonspecific interactions, as well as switching from specific or nonspecific interactions to not interacting with the DNA at all. Importantly, we imagine that the facilitated diffusion process is highly dynamic.

1.5 Facilitated diffusion

Facilitated diffusion refers to the process where a DNA binding protein locates its recognition sequence on DNA (36-38) by interacting with the nonspecific DNA that surrounds the site. In most cases – especially considering this process *in vivo* – the recognition sequence represents a tiny fraction of the overall DNA in the system. For example, there are certain transcription factors in bacteria that have only one recognition site on the whole (millions of bp) genome. One might imagine that the nonspecific DNA would be a "distraction" to the process of finding the specific site; this argument is further convincing considering the homogeneity of DNA, particularly given that the difference

between cognate and not cognate sites (where, for example, binding discrimination can differ by several orders of magnitude) can be on the order of only a few atoms.

Interestingly, initial studies of DNA binding protein and DNA systems (shortly following the observation that proteins bind to DNA in the late 1960's) (39) found that the rate of association between the protein (the first object) and the region of DNA that the protein binds to (the second object) exceeds the rate possible by diffusion (40-42), as defined by the Smoluchowski equation (43). This was shown using experiments exploring the on and off kinetic constants for the lac repressor and its operator. One proposed explanation for the phenomenon by Riggs *et al.* was that the protein was able to use the DNA flanking the recognition site as a tool to "roll" or "hop" along the DNA; these movements along the nonspecific DNA would "facilitate" the diffusion process (41). While this phenomenon is overwhelming accepted now (45 years following this work), Riggs *et al.* were interestingly very critical of this (their) initial interpretation, and deemed it unlikely (41). Furthermore, they suggested the following experiment, where one would measure the association rate between the protein and the DNA as a function of the length of DNA, as a way to confirm or rule out this facilitated diffusion mechanism (41).

This experimental approach was adopted by others, and indeed revealed that facilitated diffusion was occurring. This was most clearly experimentally demonstrated by Jack *et al.* (44) for the restriction endonucleases EcoRI and by von Hippel and coworkers, again using the lac repressor (36-38). Jack *et al*. refrained from directly commenting on the details considering how facilitated diffusion might work in the context of his experiments (44). von Hippel and coworkers developed an extensive theoretical basis for the mechanisms of different translocation mechanisms (36-38). While much of this early work remains the

foundation of the facilitated diffusion studies that have followed (up until the present), further explorations of the details of these mechanisms remain ongoing, and represent the bulk of this thesis.

More currently (in the last 15 years), facilitated diffusion studies have focused on sliding and hopping. Rigurous experimentation has explored the details of these two mechanisms, probing the exent that particular proteins use either mechanism (see Chapter 2). Consequently, much of the work of Halford, Stivers, and O'Brien lends to the notion that "all proteins will use some combination of sliding and hopping mechanisms" (see Chapter 2). However, our work here and alternative literature examples suggest that other mechnans are likely contributing to the facilitated diffusion process.

1.6 References

1. Cheng, X. (1995) Structure and function of DNA methyltransferases. *Annu. Rev. Bioph. Biom. Struct.* **24**, 293-318

2. Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* **3**, 274-293

3. Laird, P. W. & Jaenisch, R. (1996) The role of DNA methylation in cancer genetics and epigenetics. *Annu. Rev. Genet.***30**, 441-464

4. Johnson, T. B. & Coghill, R. D. (1925) Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus1. *J. Am. Chem. Soc.* **47**, 2838-2844

5. Hotchkiss, R. D. (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* **175**, 315-332

6. Hurwitz, J., Gold, M. & Anders, M. (1964) The Enzymatic Methylation of Ribonucleic Acid and Deoxyribonucleic Acid III. Purification of soluble ribonucleic acid-methylating enzymes. *J. Biol. Chem.* **239**, 3462-3473

7. Klimasauskas, S., Kumar, S., Roberts, R. J., & Cheng, X. (1994). Hhal methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357-369

8. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Gene. Dev.* **16**, 6-21.

9. Diekmann, S. (1987) DNA methylation can enhance or induce DNA curvature. *EMBO J.* **6**, 4213.

10. Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K. & McCann, P. P. (1996). S-Adenosylmethionine and methylation., *FASEB J.* **10**, 471-480.

11. Hermann, A., Gowher, H. & Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell. Mol. Life Sci.* **61**, 2571-2587.

12. Reik, W., Dean, W. & Walter, J. (2001) Epigenetic reprogramming in mammalian development. *Science* **293**,1089-1093.

13. Bird, A. (1999) DNA methylation *de novo*. *Science* **286**, 2287-2288.

14. Warnecke, P. M. & Bestor, T. H. (2000) Cytosine methylation and human cancer. *Curr. Opin. Oncol.* **12**, 68-73.

15. Bird, A. (2007) Perceptions of epigenetics. *Nature* **447**, 396-398.

16. Casadesús, J. & Low, D. (2006) Epigenetic gene regulation in the bacterial world." *Microbiol. Mol. Biol.Rev.***70**, 830-856

17. Painter, R.C., Osmond, C., Gluckman, P., Hanson, M., Phillips, D. I. & Roseboom, T. J. (2008). Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. *BJOG-Int. J. Obstet. Gy.* **115**, 1243-1249

18. Holliday, R. (2005) Epigenetics: a historical overview. *Epigenetics: official journal of the DNA Methylation Society* **1**, 76-80

19. Sandler, I. & Sandler, L. (1985) A conceptual ambiguity that contributed to the neglect of Mendel's paper. *History and philosophy of the life sciences* 3-70

20. Waddington, C. H. (1939) An introduction to modern genetics. *An Introduction to Modern Genetics*

21. Waddington, C. H. (1957) The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser." *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser.* ix+-262.

22. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K., & Schapira, M. (2012). Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. drug Discov.* **11**, 384-400

23. Strahl, B. D. & Allis, D. C. (2000) The language of covalent histone modifications. *Nature* **403**, 41-45

24. Allis, D. C., Jenuwein, T., Reinberg, D. & Caparros, M.L. (2007) Epigenetics. eds

25. Arber, W. (1965) Host-controlled modification of bacteriophage. *Annual Reviews in Microbiology* **19**, 365-378

26. Murk, I., and Kobayashi, I. (2014) To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res.* **42**, 70-86

27. Marinus, M. G. & Casadesus, J. (2009) Roles of DNA adenine methylation in host–pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol. Rev.* **33**, 488-503

28. Collier, J. (2009) Epigenetic regulation of the bacterial cell cycle." *Curr. Opin. Microbiol.***12**, 722-729

29. Collier, J., McAdams, H. H. & Shapiro, L. (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle." *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17111-17116

30. Horton, J. R., Liebert, K., Bekes, M., Jeltsch, A. & Cheng, X. (2006) Structure and Substrate Recognition of the *Escherichia coli* DNA Adenine Methyltransferase. *J. Mol. Biol.* **358**, 559-570

31. Horton, J. R., Liebert, K., Hattman, S., Jeltsch, A. & Cheng, X. (2005). Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. *Cell*, **121**, 349-361

32. Coffin, S. R., & Reich, N. O. (2009) *Escherichia coli* DNA Adenine Methyltransferase: the structural basis of processive catalysis and indirect read-out." *J. Biol. Chem.* **284**, 18390-18400

33. Allan, B. W. & Reich, N. O. (1996) Targeted base stacking disruption by the Eco RI DNA methyltransferase. *Biochemistry* **35**, 14757-14762

34. Liebert, K., Hermann, A., Schlickenrieder, M., & Jeltsch, A. (2004). Stopped-flow and Mutational Analysis of Base Flipping by the *Escherichia coli* Dam DNA-(adenine-N6)-methyltransferase., *J. Mol. Biol.* **341**, 443-454

35. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R. & Kaptein, R. (2004). Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386-389

36. Berg, O. G., Winter, R. B. & von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929-6948

37. Winter, R. B., Berg, O. G. & von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli lac* repressor-operator interaction: equilibrium measurements. *Biochemistry* **20**, 6948-6960

38. Winter, R. B., Berg, O. G. & von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry* **20**, 6961-6977

39. Riggs, A. D., Bourgeois, S., Newby, R. F. & Cohn, M. (1968) DNA binding of the *lac* repressor. *J. Mol. Biol.* **34**, 365-368

40. Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970) *lac* repressor-operator interaction: I. Equilibrium studies. *J. Mol. Biol.* **48**, 67-83

41. Riggs, A. D., Bourgeois, S. & Cohn, M. (1970) The *lac* represser-operator interaction: III. Kinetic studies. *J. Mol. Biol.* **53**, 401-417

42. Adam, G. & Delbrück, M. (1968) Reduction of dimensionality in biological diffusion processes." *Structural chemistry and molecular biology.* 198

43. Smoluchowski, M. von. (1917) Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z. phys. Chem* 92, 9

44. Jack, W. E., Terry, B. J. & Modrich, P. (1982) Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4010-4014

II. DNA Looping Provides for "Intersegmental Hopping" by Proteins: a Mechanism for Long-Range Site Localization

Parts or sections of this chapter were taken with permission from: Pollak, A. J, Chin, A. T., Brown, F. L. H. & Reich, N.O. (2014) DNA looping provides for "intersegmental hopping" by proteins: a mechanism for long-range site localization. *J. Mol. Biol.* DOI: 10.1016/j.jmb.2014.08.002

2.1 Abstract

Studies of how transcription factors and DNA modifying enzymes passively locate specific sites on DNA have yet to be reconciled with a sufficient set of mechanisms that can adequately account for the efficiency and speed of this process. This is especially true when considering that these DNA binding/modifying proteins have diverse levels of both cellular copy numbers and genomic recognition site densities. The monomeric bacterial DNA adenine methyltransferase (Dam) is responsible for the rapid methylation of the entire chromosome (with only ~100 Dam copies per cell) and the regulated methylation of closely spaced sites, which controls the expression of virulence genes in several human pathogens. Provocatively, we find Dam travels between its recognition sites most efficiently when those sites are ~500 base pairs apart. We propose that this is manifested by Dam moving between distal regions on the same DNA molecule, which is mediated by DNA looping, a phenomenon we designate as intersegmental hopping. Importantly, an intermediate found in other systems including two simultaneously bound, looped DNA strands is not involved here. Our results suggest that intersegmental hopping contributes to enzymatic processivity (multiple modifications), invoking recent reports that demonstrate DNA looping can assist in site finding. Intersegmental hopping is possibly used by other sequence specific DNA

binding proteins, such as transcription factors and regulatory proteins, given certain biological context. While a general form of this mechanism is proposed by many research groups, our consideration of DNA looping in the context of processive catalysis provides new mechanistic insights and distinctions.

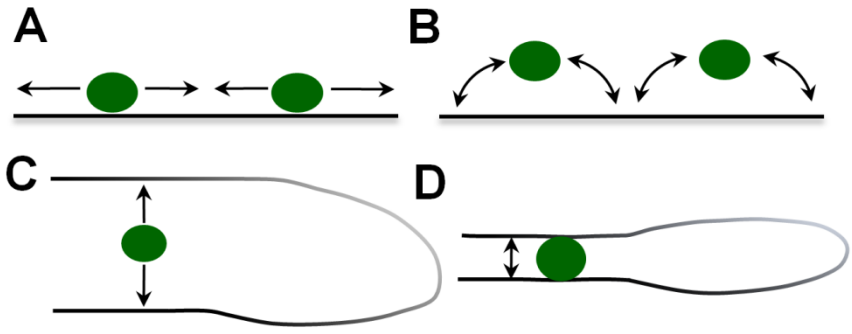
2.2 Introduction

Site specific DNA binding proteins, such as transcription factors, DNA methyltransferases, restriction endonucleases (ENases), and DNA repair enzymes locate their recognition sites with remarkable speed and efficiency. While non-specific DNA surrounding these sites aids in the search (1), the mechanism(s) of how proteins move passively (without ATP) and rapidly along nonspecific DNA remain incompletely understood. Interestingly, this field of facilitated diffusion has a rich history, dating back to pioneering works of Delbrück in 1968 and Riggs in 1970 (see Chapter 1). In short, DNA binding proteins are overwhelmingly likely to associate with non specific DNA during the initial encounter of the protein and the DNA, particularly under *in vivo* conditions where sites are usually quite rare. What follows is a series of kinetic steps where the protein samples different regions of the DNA in search of its sites while moving along the DNA. The presence of nonspecific DNA has been broadly accepted to assist in the association between the protein and the segment of DNA containing the site; interestingly, one could imagine that the nonspecific DNA would impair this process, distracting the protein with the overwhelming presence of nonproductive sites. What remains incompletely understood is the precise nature of the movements along the DNA.

In particular, current translocation models that largely invoke sliding and hopping processes, which are based on extensive *in vitro* (2) and recent *in vivo* studies (3), are

inadequate to solely account for searches over thousands and hundreds of thousands of base pairs routinely achieved by these proteins. With sliding (4,5), the protein maintains continuous contact with DNA, like a roller coaster car on a track (Figure 2.1A). Each movement is a single base pair (bp) step in either direction, likely constraining each protein to short stretches of DNA (<100 bp). Hopping involves a rapid series of dissociation and reassociation events, were the protein remains loosely associated with the DNA (Figure 2.1B) (6,7,8). Details about the average trajectory of each hop, however, remain incompletely characterized. In general, hopping allows for larger movements along DNA than for sliding, but neither mechanism has been shown to contribute to kb searches. Sliding/hopping mechanisms are likely coupled with other, presently ambiguously described long-range 3-D movements (9), which we attempt to further elucidate here.

Figure 2.1. Translocation models: A) Sliding involves a close association between the protein and the DNA. B) Hopping includes small jumps. C) Intersegmental hopping (proposed here) combines aspects of hopping and intersegmental transfer for fluid, long-range protein translocation. D) A ternary complex is obligatory for proteins to move between distal DNA segments for intersegmental transfer.

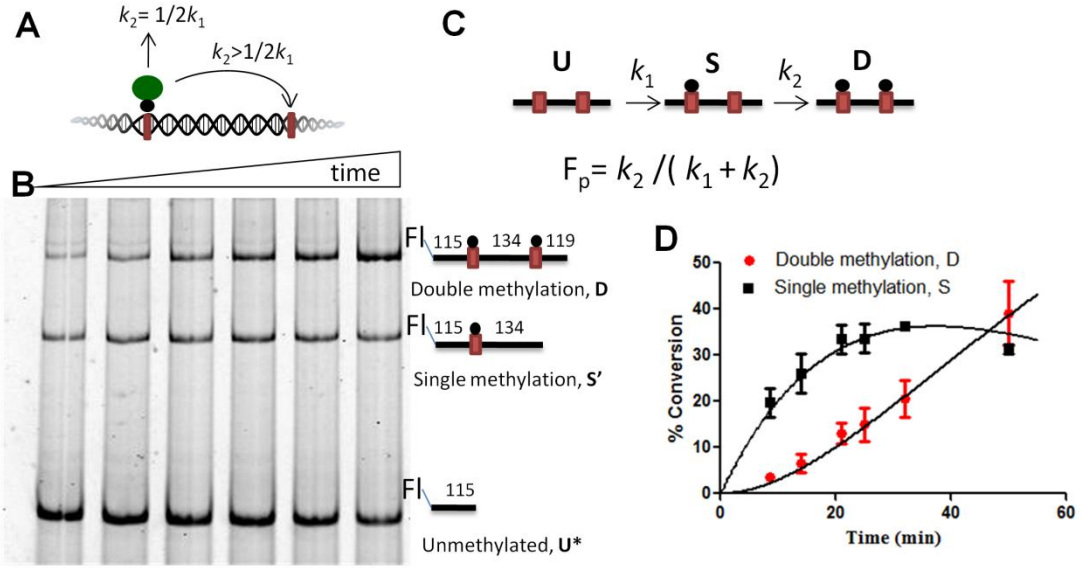


The highly processive *Escherichia coli* DNA adenine methyltransferase (Dam) is an ideal model system to study long range translocation mechanisms: relatively few enzymes

(~100 Dam enzymes per cell) methylate ~20,000 recognition sites (5'-GATC-3') at the N^6 position of adenine in approximately 20 minutes, an essential component of mismatch repair (10). Dam is interestingly involved in several other processes in the cell, including epigenetic gene regulation (10,11). Other low copy proteins (e.g., transcription factors) also must transverse long segments of the genome to locate their site of action (discussion). Here we use an activity-based assay and a series of two site DNA substrates to study processivity (Fig. 2.2) to deduce translocation mechanisms. Processivity (F_p) is quantified as the probability that Dam methylates both sites within a single binding event between the enzyme and the DNA.

Changes in processivity with increases in inter-site distances are commonly used, and are a powerful tool to deduce translocation mechanisms and the details concerning them (2,12,13). With sliding, processivity will decrease more sharply with increasing intersite distances than for hopping (12,14) (below). While the translocation kinetics of some proteins are well described by the sliding and/or hopping processivity equations (2), several reports contain data that cannot be reconciled with these models (see discussion). Both mechanisms involve movement following the contour of the helix and predict that processivity should decrease as inter-site separation is increased; this decrease is less dramatic for hopping.

Figure 2.2. Processivity assay. **A**) Dam (green circle) immediately following the first methylation (black circle) at a GATC site (bold red lines). Processivity is based on the probability of a second methylation event following an initial one during a single binding event. **B**) The reaction includes 7 nM Dam, 400 nM DNA (symmetric substrate, fluorescein labeled), 30 μM SAM (S-adenosyl methionine), 30 mM NaCl, and buffer at 37 degrees. Heat quenched samples were digested with DpnII endonuclease (cuts unmethylated GATC sites), and reaction products were separated by non-denaturing PAGE and imaged using a typhoon phosphoimager. **C**) The relative amount of each species of the reaction is fit to a sequential reaction mechanism model to derive the rate constants k_1 and k_2, and ultimately the processivity value (F_p) (Equations 4-6). **D)** Fit of data to model for substrate with 134 base pairs between sites (5 trials, mean and standard deviation are plotted, Substrate 4B-2, Table 2.1).



Several lines of evidence form the basis for distinguishing hopping from sliding. While the sliding and hopping models predict how processivity will vary as a function of the spacings between the two sites, each model contains parameters that can vary widely (see chapter 3), making the distinction between the two mechanisms difficult to prove (1,2,7,8). For example *a* (from the hopping model) was experimentally determined by Halford and colleagues to be ~56 bp (12), while Stivers and colleagues determined the value to be ~ 10 bp (13). This makes sense because thos different proteins will behave differently, but this

difference shows that this analysis can be potentially inconclusive, especially considering how, to the best of our knowledge, *a* has only be determined in these two systems. The use of processivity is particularly difficult to distinguish between related mechanisms when proteins use a combination of hopping and sliding, which seems to be the case for most proteins. However, clever modulations to the standard processivity analysis move towards demonstrating distinctions between hopping and sliding. One is the use of salt (see below and chapter 3) where the processivity of a hopping protein is predicted to be impacted severely by salt. Another approach utilized the positioning of non-palindromic recognition sites on both on the same and opposite strands of the DNA (see Discussion). The enzyme was more processive when the sites were on the same DNA stand with short spacings between the sites (indicative of sliding); however when the sites were spaced farther part, processivity was equal when the sites were on the same or opposite DNA strand (indicative of hopping). Another approach relies on incubating a non-catalytic, site specific protein in between the two sites (see chapter 4). The retention of some level of processivity is indicative of a hopping event. A final approach discussed here involved the use of a weakly bound, small molecule inhibitor in the background during the processivity analysis (see discussion). During hopping events, a population of the enzyme will be captured by the inhibitor. The concentration dependence of processivity modulations formed the basis for an estimated hopping distance of ~7 nm (8).

Importantly, long distance translocation (>1 kb) is extremely unlikely solely considering sliding/hopping mechanisms (2). Data presented here and prior literature reports (see discussion) suggest an alternative translocation mechanism exists where individual proteins can scan large regions of the genome. We propose that intersegmental hopping, involving

the movement of proteins between looping segments of DNA, potentially presents an artful solution to the extensive site location undertaking (Figure 2.1C) (below). (This is distinct from the well-accepted intersegmental transfer mechanism (Figure 2.1D), where a ternary complex is required (14)). While similar components of intersegmental hopping are heavily acknowledged (discussion), critical mechanistic details remain uncharacterized. Importantly, here we consider DNA conformation (15) in relation to our processivity analysis. Specifically, the looping of the DNA to bring either specific or non-specific segments into proximity is explored as an avenue for efficient protein translocation.

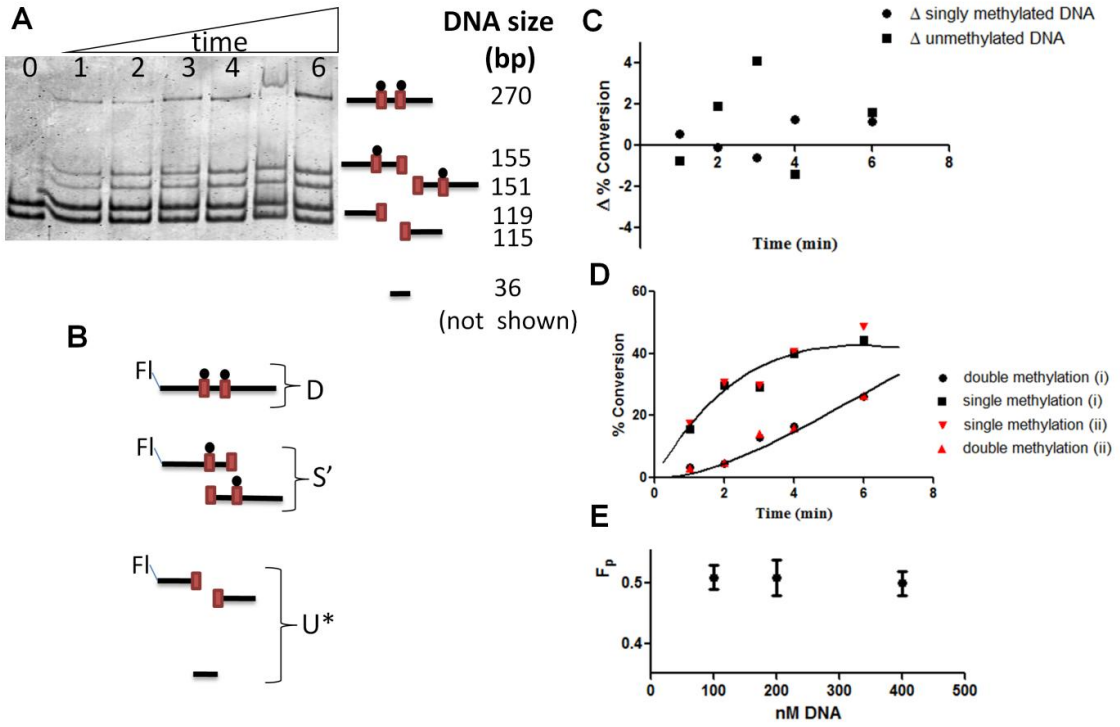
2.3 Results

2.3.1 Processivity Assay

Processivity (F_p) is based on the probability that Dam methylates the second site after methylating the first site on a two site DNA substrate within a single binding event between the enzyme and the DNA (Figure 2.2). Importantly, DNA is in excess of enzyme (400:7) to limit situations where two enzymes are bound to one molecule of DNA. While rare, situations will arise where two enzymes are bound to a single DNA substrate at one time, potentially obscuring the processivity analysis. However, we maintain the value of the processivity analysis lies in its comparative nature over a series of substrates, where this potential problem will equally arise for all substrates. Therefore, our interpretation of the processivity trends endures in spite of this concern. At each time point during the methylation reaction, the amount of unmethylated, singly methylated, and doubly methylated DNA is separately quantified using a methylation sensitive restriction endonuclease based readout (DpnII cuts unmethylated GATC sites following heat quench and slow cooling (16)) accompanied by a PAGE gel analysis (Figure 2.2B). The substrate is

fluorescein labeled on one end to simplify the gel analysis.

Figure 2.3. Processivity assay controls. **A**) Processivity assay using substrate 4B-1 without a fluorescein label (Sybr Au staining). Gel shows all DNA cutting fragments. **B**) Schematic of DNA fragments from the processivity assay. Since the substrates are fluorescein labeled on one end, "ghost bands (unlabeled) for single and unmethylated species are not visualized. **C**) Difference between the DNA density for singly (black circles) and unmethylated DNA (black squares) at each time point from **A**. Minimal differential confirms no site preference. **D**) Fit of unlabeled substrate to sequential methylation model using either the bands that would be labeled (i) or the bands that would be "ghost bands (ii)" from **B**, both resulting in a processivity value (F_p = 0.42) consistent with the value obtained by using the fluorescein labeled substrate. **E**) Processivity values for substrate 4B-5 as a function of DNA concentration. Ratio of enzyme to DNA remains constant. The lack of response in processivity shows that at 400 nM, Processivity is not obscured by different DNA strands interacting with each other.



The percent conversion of single and double methylation events are calculated using the total density within each individual lane. D is the top band, corresponding to doubly methylated DNA. S' is the middle band, corresponding with singly methylated DNA, and S'

needs to be doubled to account for the other singly methylated DNA fragment in the "ghost band." Finally, U* is the bottom band, and the amount of unmethylated DNA can be accounted for by subtracting S' from U* (Figures 2.2B, 2.3B). The uncut band is not doubled as multiple fragments represent one species. The amount of each species can be calculated with a fluorescein label on only one end of the DNA (Equations 1-3).

Total density per lane: $T = D+2S'+U^*-S'$ (**1**)

$T = D + S' + U^*$ (**1a**)

Singly methylated = $2S'/T$ (**2**)

Doubly methylated = D/T (**3**)

This approach is appropriate because all the substrates are essentially symmetric, with no catalytic preference for either site. This is confirmed by a processivity assay using an unlabeled substrate with a staining procedure where all DNA fragments can be visualized (Figure 2.3A,C,D). Each single and each unmethylated band pairs are nearly equal at each time point, showing no preference at either site (Figure 2.3C). The flanking DNA immediately surrounding the sites is the same for all substrates in this study. Importantly, our interest is in how processivity changes with different DNA lengths.

Another control to validate our assay concerns DNA concentration. At 400nM, one could hypothesize that separate DNA strands might become proximal, possibly causing the enzyme to switch between those strands, which would introduce a problematic scenario where high DNA concentration influences processivity. We sought to empirically rule out this scenario by titrating the amount of DNA while keeping the ratio of DNA to enzyme constant. If the DNA concentration is high enough such that different strands coming together influences the enzyme, processivity will then decrease as a function of increasing

DNA. Therefore, we preformed our processivity assay at DNA concentrations of 100nM, 200nM and 400nM, with commensurate changes in enzyme concentration (the latter is our standard assay level) (Figure 2.3E). As processivity does not decrease through this series, separate DNA strands are unlikely to interact at 400nM DNA.

The accumulation of each species (Equations 1-3) during multiple turnovers is fit to a sequential reaction scheme (17) (Equations 4 and 5, Figure 2.2C), where k_1 corresponds to the 1st order rate constant from unmethylated to singly methylated DNA, and k_2 corresponds to the 1st order rate constant from singly to doubly methylated DNA. The relative amount of each species from the gel densitometry at each time point is the basis for the k_1 and k_2 values (Figure 2.2D).

$$\text{Singly methylated (S)} = \frac{k_1}{(k_2-k_1)} \times (e^{-tk_1} - e^{-tk_2}) \qquad (\mathbf{4})$$

$$\text{Doubly methylated (D)} = 1 + \frac{1}{(k_1-k_2)} \times (k_2 e^{-tk_1} - k_1 e^{-tk_2}) \qquad (\mathbf{5})$$

Processivity (the relevant outcome of the analysis) relates the values of k_2 and k_1 (Equation 6).

$$F_p = \frac{k_2}{k_1+k_2} \qquad (\mathbf{6})$$

Importantly, the first order process involving the searching of the DNA for the site(s) following the association between the enzyme and the DNA is rate limiting. Note the rate constants decrease with increases in the size of the DNA, consistent with this notion (Table 2.1). Also, if the association between the enzyme and DNA were rate limiting, then Figure 2.3E would show modulations in processivity with changes in enzyme and DNA concentrations, which is not the case (18). k_2 involves searches on DNA with one available site whereas there are two available sites involved in searching for k_1. Therefore, the lower

limit of k_2 is expected to be half of k_1 (18), meaning the range of processivity values from lowest to highest is between 1/3 and 1 (Equation 6). For example, when $k_2 = k_1/2$ the enzyme is unprocessive (F_p = 1/3) as each methylation event involves the enzyme starting from solution to find its recognition site, resulting in a large accumulation of singly methylated species before the production of doubly methylated species. We demonstrate the lower limit of processivity detection with an experiment introducing 230mM salt into our processivity reaction. This high salt level lowers processivity (see below), and substrate 4B-1's processivity under these salt conditions is $F_p = 0.32 \pm .02$ (data not shown). Overall, this general processivity method has been previously used (18,19), is similar to a method based on the relative accumulation of singly-modified intermediates (20), and is extensively justified here.

2.3.2 Processivity data suggestive of DNA looping

Results with Dam unexpectedly showed increases in processivity as the intersite distances increase from 36 bp to 605 bp, followed by a downturn in processivity from 605 bp to 798 bp (Figure 2.4, Table 2.1). These processivity trends are not predicted by sliding or hopping models of protein translocation. Moreover, such an increase to the best of our knowledge has never been reported. Interestingly, our trends are reminiscent of the length dependence of DNA looping (Figure 2.4C) (21-25). The looping of DNA is limited by its persistence length (L) and distance between looping sites (s); short DNA segments (<0.15 kb) are too rigid to fold back on themselves and long segments (>1 kb) are entropically unlikely to find a loop forming configuration (26). These effects are well quantified by (22)

$$J_m(s) = \left(\frac{3}{4\pi sL}\right)^{3/2} \times e^{\left(\frac{-8L^2}{s^2}\right)} \times (46.22 Mbp^3) \qquad (7),$$

which expresses the effective local concentration of one site around the other at equilibrium. (The conversion factor to molar units given in Equation **7** presumes the lengths *s* and *L* are measured in bp; 1bp = 0.33nm.) The probability of loop formation sharply increases between 200 and 500 bp, and then gradually decreases. This equation is appropriate for both co-localization of DNA ends (21) and co-localization of two sites in the middle of a DNA polymer (22), the latter of which is relevant to this study. We stress that Equation 7 reflects the effective local concentration of site two in the vicinity of site one *irrespective of the orientation of the two sites relative to one another*. The $J_m(s)$ used in this study are thus distinct from, and not to be confused with, "j factors" used in the modeling of DNA ligation experiments (24,25), which require directional and torsional alignment of the terminal DNA bases in addition to spatial co-localization predicted by Equation 7. We also note that Equation 7 closely approximates Equation 50 of Shimada and Yamakawa (23) over the values of *s* relevant to this study, however eq. 7 has the advantage of being applicable over arbitrarily large site separations and was adopted for its generality in this regard.

Extensive prior reports form the basis for this DNA looping equation. Initial work by Baldwin *et al.* in 1981 explored the probability of loop formation as a function of the total size of the DNA (21). They employed a strategy where the DNA's (of various sizes) have sticky ends corresponding to EcoRI ENase restriction fragments. Upon the addition of ATP and DNA ligase, the DNA ends are joined. One possible outcome is distinct DNA fragments will be joined together, forming multimers. Another possibility involves the sticky ends from a single DNA fragment joining together, generating a DNA loop. The ability to define the probability of loop formation for a single DNA length relies on discriminating between the looped DNA and the inter-DNA joined multimers. This Jm(s) is experimentally

determined using PAGE, where circular DNA's migrate differently than linear ones of the same molecular weight. This procedure is repeated for DNA's of variable length. The ratio of the two species is the basis of J_m(s). This auto-ligation procedure has been used extensively to explore the torsinal properties of DNA (see below).

Figure 2.4. Processivity trends invoke DNA looping, are inconsistent with sliding/hopping models. **A**) Schematic of DNA substrates. The distance between sites changes; the DNA flanking the sites are held constant at 115 and 119 base pairs. **B**) Processivity increases with greater inter-site distances (Table 2.1, substrates 4B-1-10). Modeling (red) from Equation 12 heavily suggests the influence of site-specific intersegmental hopping, although, deviations from the modeling are anticipated. **C**) The probability of loop formation (L=150 bp) as a function of the distance between recombination sites (22) (black line). This trace is similar to our processivity data in **A**. Processivity trace for a sliding enzyme (12) (blue line) is dramatically inconsistent with our processivity trends for Dam. **D**) Processivity data for substrate 4B-10, which has 798 bp between sites (3 trials, mean and standard deviation are plotted).To the right is the corresponding PAGE gel for substrate 4B-10. See gel in Figure 2.2B for details on annotation. A theoretical trace for a completely unprocessive reaction is shown, with k_1= 0.019 min^{-1} and k_2= 0.0095 min^{-1}.The single methylation trace is depicted by dashes and double methylation trace is depicted by dots. The processivity trace for substrate 4B-10 has a dramatically smaller accumulation of singly methylated product than the theoretically unprocessive example.

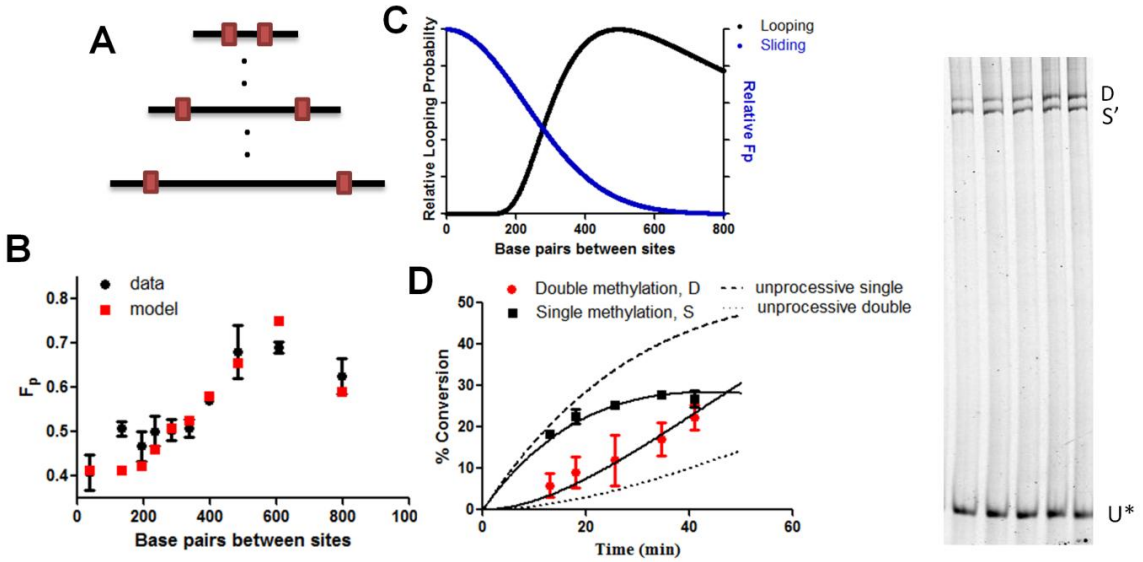


Table 2.1. Dam and EcoRI ENase processivity data (Figure 2.4B, Figure 2.8C): Substrates with "E" in the name are for EcoRI ENase substrates, all others are for Dam. Dam error is calculated from replicates of 2 or more individual experiments. EcoRI ENase error is calculated from replicates of 4 or more individual experiments. Error for F_p is calculated using F_p values generated from individual experiments.

Substrate name	Total size (bp)	Distance between sites (bp)	k_1 (min^{-1})	k_2 (min^{-1})	F_p
4B-1	270	36	0.136 ± 0.028	0.092 ± 0.004	0.41 ± 0.04
4B-2	368	134	0.026 ± 0.003	0.027± 0.003	0.51 ± 0.02
4B-3	428	194	0.038 ± 0.005	0.033± 0.0002	0.47 ± 0.03
4B-4	468	234	0.026 ± 0.003	0.026± 0.003	0.50 ± 0.03
4B-5	518	284	0.026 ± 0.002	0.026 ± 0.001	0.50 ± 0.02
4B-6	569	335	0.031 ± 0.005	0.032 ± 0.005	0.51 ± 0.02
4B-7	629	395	0.023 ± 0.004	0.030 ± 0.005	0.57 ± 0.01
4B-8	716	482	0.014 ± 0.003	0.030± 0.004	0.68 ± 0.06
4B-9	839	605	0.007 ± 0.002	0.015 ± 0.004	0.69 ± 0.01
4B-10	1032	798	0.017 ± 0.001	0.029 ± 0.006	0.63 ± 0.04
4C-1	66	36	0.225 ± 0.018	0.112 ± 0.015	0.33 ± 0.04
4C-2	110	36	0.169 ± 0.007	0.092 ± 0.004	0.35 ± 0.02
4C-3	636	36	0.138 ± 0.023	0.112 ± 0.004	0.45 ± 0.03
5CE-1	116	42	0.019 ± 0.004	0.026 ± 0.007	0.58 ± 0.06
5CE-2	276	42	0.016 ± 0.002	0.022 ± 0.005	0.58 ± 0.05
5CE-3	642	42	0.008 ± 0.001	0.010 ± 0.002	0.58 ± 0.05

2.3.3 Modeling

Based on the above observations, we suggest that Dam may be capable of "specific intersegmental hopping", whereby the enzyme directly transits between binding sites (and/or regions adjacent to the binding sites) when the sites are brought in close proximity due to the looping of the DNA chain. The contribution of specific intersegmental hopping to F_p, as predicted by Equation **7**, can be captured within an elementary kinetic model. Processivity quantifies the enhancement of k_2 over $k_1/2$ and we imagine that the observed k_2 values originate from three competing kinetic mechanisms: processive methylation of the second site by the same Dam as the first site via mechanisms other than specific intersegmental

hopping (including, but not limited to: sliding, hopping and non-specific intersegmental hopping) with rate constant k_{non}, processive methylation via specific intersegmental hopping by Dam with rate constant k_{sp}, and methylation of the second site by a new Dam from solution with rate constant $k_1/2$.

$$k_2 = k_1/2 + k_{non} + k_{sp} \quad (\mathbf{8}).$$

Non-specific intersegmental hopping (below) refers to looping enabled hops (which can occur anywhere along the DNA) that do not involve direct site-to-site transfer. Equation 7 suggests that

$$k_{sp} = k' J_m \quad (\mathbf{9})$$

where k' is a second order rate constant associated with specific intersegmental hopping (Equation 9). The first order rate constant k_{non}, would be expected to be a potentially complex function of total DNA length as well as the positions of the recognition sites along the DNA contour. To simplify the analysis, we introduce the ansatz that

$$k_{\text{non}} \approx (\alpha - 1/2) k_1 \quad (\mathbf{10}),$$

where α is a dimensionless constant defined such that α=1/2 corresponds to a system that would be non-processive if specific intersegmental hopping were artificially turned off (Equation 10). α>1/2 indicates an enzyme that would be processive even if specific intersegmental hopping was unavailable. This assumption leads to

$$k_2 = \alpha k_1 + k' J_m \quad (\mathbf{11}).$$

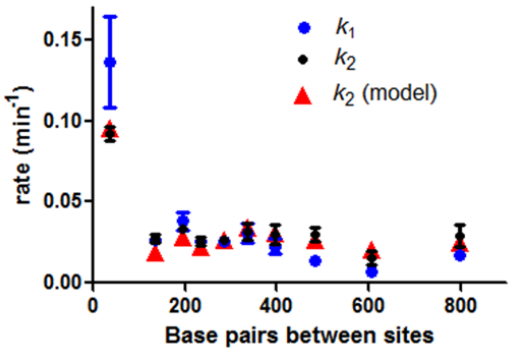
Substituting Equation 11 into Equation 6 leads to Equation 12

$$F_p(s) = \frac{\alpha k_1(s) + k' J_m(s)}{(1+\alpha) k_1(s) + k' J_m(s)} \quad (\mathbf{12}),$$

where the dimensionless constant α captures the effects of processivity distinct from specific intersegmental hopping (F_p= α/[1+ α] absent specific intersegmental hopping, i.e. α>1/2 indicates an enzyme that would be processive even if specific intersegmental hopping were turned off) and k' is the second order rate constant associated with specific intersegmental hopping.

The experimental data is fit well by Equation 12 (Figure 2.4B). The best fit values of α=0.70 and k' = 1.3 x 10^5 M^{-1} min^{-1} were determined via linear least squares fitting, assuming L=150 bp for the persistence length of DNA (12,21) in Equation **7** for J_m. (The fitting was also repeated using other values of the persistence length and it was found that the best possible fit was obtained with L=144 bp, however the improvement over the L=150 bp case is nearly invisible in Figure 2.4B and is not plotted.) Using the experimental values for k_1 and the best-fit values of α and k' provided above, the processivity values are plotted for each value of the site separation; these are the "model" points displayed in Figure 2.4B. Figure 2.5 displays a comparison between experimentally determined k_2 values and k_2 values predicted using the best fit terms α and k' as a function of k_1 via Equation 11.

Figure 2.5. The k_1 and k_2 experimental data, and k_2 derived from the model segregating intersegmental hopping from other translocation mechanisms (Equation 11, Figure 2.4B) are plotted.



The good agreement between Equation 12 and the experimental data serves as compelling evidence in favor of the specific intersegmental hopping mechanism for Dam. However, it should be stressed that the finding of $\alpha = 0.70$ indicates that other mechanisms do contribute to Dam processivity, in addition to specific intersegmental hopping. Further experiments (below) suggest non-specific intersegmental hopping certainly contributes to the magnitude of α. The present model is insufficiently detailed to speculate on the nature of the contribution from each mechanism to α, and assumes that their contributions are independent of the site spacing, at least for the spacings considered here. We stress that the modeling employed in this work is clearly of a minimalist nature and was developed to assess the importance of specific intersegmental hopping without needing to elucidate the full details of all mechanisms contributing to processive turnover. A more elaborate model, both from the standpoints of the underlying kinetic scheme and of the individual mechanisms contributing to rate constants, would be needed to fully explain the observed processivity in Dam and the relative contributions from all competing mechanisms.

2.3.4 Intersegmental hopping does not involve a bridging complex

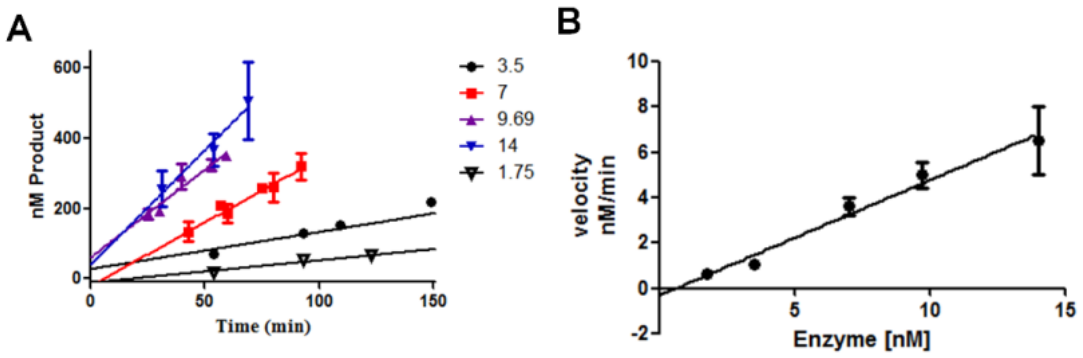
If looping contributes to the rapid search kinetics of Dam, the enzyme is able to move from one segment of the same DNA molecule to another without following the contour of the DNA (sliding/hopping). Similar intersegmental transfer mechanisms have been invoked for enzymes like the restriction endonucleases SfiI and FokI (27,28), which oligomerize and simultaneously bind to two separate regions of the DNA (forming a bridging complex). The involvement of looping by these enzymes was confirmed based on rigorous studies using an equation similar to our Equation **7** (29).

For example, the FokI ENase was shown to dimerize and simultaneously cut two sites on a single DNA fragment, intricately depending on the distance between the two sites over a range where the sensitivity of the torsinal properties of DNA can be revealed (28). Specifically, the cutting velocity displayed phase dependent changes between 170-200 bp. When the two sites were able to face each other the cutting velocity was enhanced, and when they were out of phase, the cutting was diminished. This oscillation had a periodicity of ~10bp, corresponding with the periodicity of DNA. This oscilliatory activity trend was only observed between 170-200 bp, as shorter intersites spacings correspond to undetectable loop formation probabilities and thus low activity; longer intersite spacings correspond to a (non oscillatory) increase in activity corresponding to increases in DNA distance. The FokI ENase is not alone in the necessity of such a loop formation to form to allow cutting. This oscillatory readout was also found with the SfiI ENase (27). This mechanism is also relevant to other ENases that recognize 8 bp sequences, and this loop formation requirement may be a general cutting mechanism for these types of ENases (53). These examples provide precedence for the contribution of DNA looping towards the related intersegmental transfer mechanism.

Evidence for the oligomerization component of intersegmental transfer relied on the titration of enzyme resulting in a nonlinear dependence on the catalytic rate: as the concentration of the enzyme reached a level to support the necessary oligomeric state to form the loop complex, the rate of catalysis was enhanced. Unlike SfiI, Dam is a monomer under our experimental conditions (30) (and crystallographically (31)), and we further demonstrate this based on a similar kinetic approach. Dam's rate of methylation increases

linearly with enzyme concentration (unlike SfiI), suggesting that Dam remains a monomer under our experimental conditions (Figure 2.6).

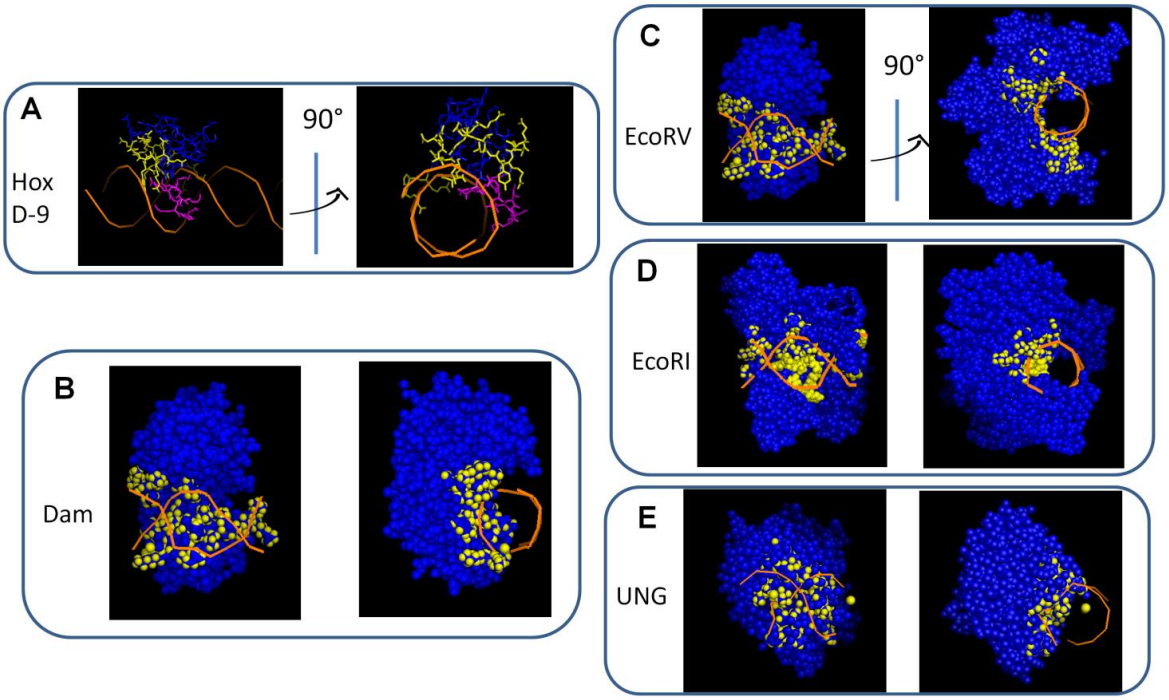
Figure 2.6. Methylated product formation of a substrate shown to undergo intersegmental hopping mediated by DNA looping (substrate 4B-9, 605 bp between the two sites) was monitored using a radiolabelled cofactor (tritiated SAM) based assay. A) Steady state assays with 400 nM DNA, 30μM radiolabeled SAM, and 1.75-14nM Dam enzyme. Experiments were done in 2-5 replicates, mean and S.D. are shown. Here, total methylation is monitored, with no information distinguishing methylation at individual sites. B) The velocity is graphed as a function of enzyme concentration. As expected, over a ten-fold range of enzyme concentration, encompassing what is used in our processivity assay (7 nM), the rate of methylation increases linearly with increasing enzyme. This suggests that Dam remains a monomer, making it unlikely that it oligomerizes to facilitate a bridging loop complex.



Intersegmental transfer can also involve bridging by two distinct DNA binding sites on a single protein. For example, extensive experimental and computation analysis suggests that HoxD-9 transfers between DNA segments of using its disordered, positively charged tail (32-35). However, this mechanism is unlikely for Dam. This conclusion is based in part on a computational model, DISPLAR, that predicts residues involved in nucleic acid binding (36) (Figure 2.7). For Dam, no alternative DNA binding residues were identified outside of the protein's known DNA binding cleft (Figure 2.7B), and it does not have a disordered tail. Furthermore, DISPLAR applied to the EcoRI and EcoRV endonucleases (Figure 2.7) both of which can use intersegmental hopping similarly to Dam (see discussion) (37,38,39), also

show no alternative DNA binding regions. Finally, AAG glycosylase undergoes intersegmental hopping without the use of its positively charged disordered tail (40), making this a less general mechanism. As intersegmental bridging by two separate DNA strands is unlikely for Dam, we carried out a systematic investigation to further determine if Dam utilizes intersegmental hopping.

Figure 2.7. DNA binding residues as identified by DISPLAR (36). Using crystal structures, DISPLAR incorporates the concentration of positively charged residues, solvent accessibility and sequence conservation as the criterion for predicting which residues are likely to be involved in DNA binding. . Importantly, DISPLAR is much more apt to incorrectly identify residues as capable of DNA binding, than it is to overlook any potential nucleic acid binding residues. Protein is blue, DNA is an orange ribbon, and DISPLAR identified DNA binding residues are yellow. All residues identified are within known DNA binding regions. **A**) Hox D-9; PDB ID 1HDD. DISPLAR correctly identifies the second DNA binding region on HoxD-9's N-terminal chain, empirically validating the model. Purple represents this secondary DNA binding region. **B**) DAM; PDB ID: 2G1P. This suggests that a scenario where Dam simultaneously binds two DNA strands is unlikely. **C**) EcoRV ENase; PDB ID 2RVE **D**) EcoRI ENase; PDB ID 1ERI. **E**) UNG; PDB ID 2OYT. UNG utilizes sliding and hopping and also can only bind DNA in its expected cleft (13).



2.3.5 Total DNA length affects processivity

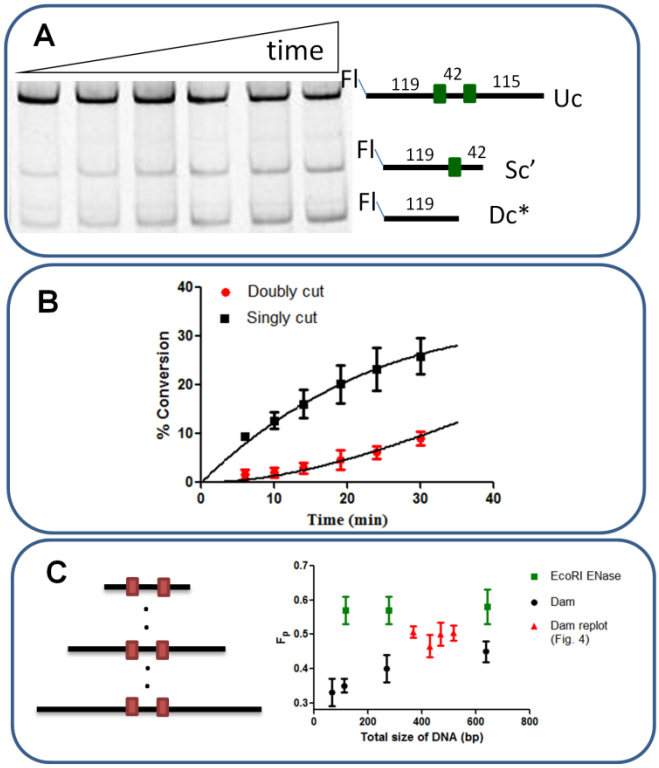
We first sought to test if the processivity is altered when the distance between the two sites is kept constant as the total length of the DNA is increased. The sliding/hopping models predict that modulations to the amount of DNA surrounding the two sites should have no effect on processivity. Interestingly, this assumption has yet to be challenged for other proteins thought to rely on sliding/hopping mechanisms. We first tested this by employing our processivity analysis to the EcoRI ENase (Figure 2.8A,B) (18), which relies heavily on a sliding mechanism (41,42) on substrates with a 42 bp intersite distance. Here, cutting events are directly monitored where a slight variation from our Dam procedure (above) is used to determine the relative amounts of singly and doubly modified species (Equations 13-15). Uc represents uncut, Sc' represents singly cut once doubled, and Dc* represents doubly cut once the singly cut fragments are subtracted.

Total density per lane: $T = Uc+Sc'+Dc^*$ (**13**)

Singly cut = $2Sc'/T$ (**14**)

Doubly cut = $(Dc^*-Sc')/T$ (**15**)

Figure 2.8. Non-specific intersegmental hopping affects processivity for Dam, but not for EcoRI ENase. **A**) Processivity assay using EcoRI ENase, an enzyme known to utilize sliding. 26 nM EcoRI ENase was added to 400nM DNA, 5mM $MgCl_2$, and buffer at 37 degrees. Uc represents uncut DNA, Sc' represents singly cut DNA once doubled, and Dc* represents doubly cut DNA once the singly cut fragments are subtracted (see text). **B**) Fit of EcoRI ENase assay (from **A**) to the processivity model for Substrate 5CE-2 (Table 2.1). **C**) Flanking DNA increases processivity with substrates having identical inter-site distances of 36 bp for Dam (black circles; Table 2.1, substrates 5C-1-3). Processivity increase scales with the probability of loop formation which is based on the total size of the substrates. Mean and of 3 trials are plotted. EcoRI ENase (green circles; Table 2.1, substrates 5CE-1-3) (42 bps between sites) is unaffected by flanking DNA as intersegmental hopping does not contribute to its processivity. Mean and S.D. of 4 trials are plotted. Substrates unlikely to undergo site-specific intersegmental hopping by Dam (red, replot from Figure 2.4B) can utilize non-specific intersegmental hopping.



These values can then be inserted into Equations 4-6 to determine F_p. For ENases, once the single cut is made the enzyme has a 50% chance to be on the cleaved DNA fragment with the second site (12) (Dam, in contrast, does not alter the DNA size following catalysis).

This makes F_p=0.66 the highest possible processivity value for EcoRI ENase (the general range is 0.33-1, see above). In agreement with previous reports, EcoRI ENase is highly processive with small intersite spacings (42bp) (42) (the ionic strength explains the slight decrease from the theoretical maximum). The robustness of this EcoRI ENase assay is demonstrated as processivity is lowered with larger intersite spacings and higher ionic strength (see chapter 3). Importantly, as expected, processivity is insensitive to changes in the length of DNA surrounding the sites and the total substrate size (Figure 2.8C).

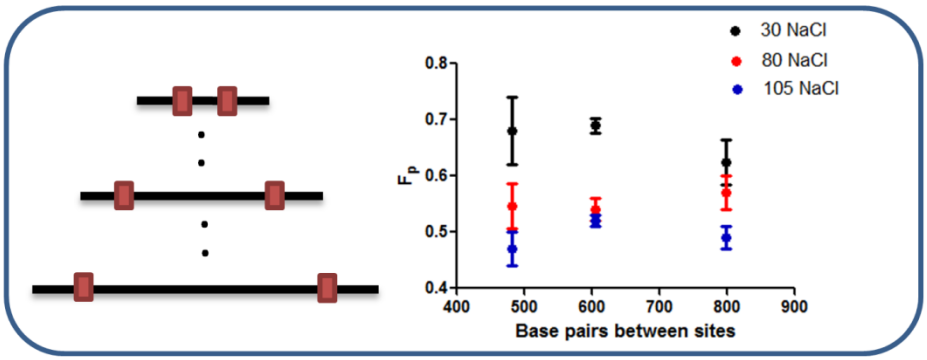
In contrast, Dam's processivity increases as a function of total DNA size with 36 bp between sites (Figure 2.8C). This is consistent with an intersegmental hopping mechanism since the DNA is increasingly more likely to loop back on itself within the total length regime of this data set (21). Notably, this non-specific looping is considered as a function of the total length of the DNA, not the distance between sites as in Fig. 2.4B. Looped segments that become proximal certainly include non-specific DNA. Thus, processivity should increase up to ~500 bp total substrate size (21) and then plateau (looping will not decrease in this case with large substrates), which is what is observed (Figure 2.8C). A replot of substrates from Figure 2.4B that are unlikely to undergo site-specific intersegmental hopping (encompassing those with 134 to 284 bp between the two sites), but are large enough to support non-specific intersegmental hopping further supports this. Importantly, a comparison between substrate 4B-7, which is overall shorter yet has a longer site to site distance and is therefore more processive than substrate 5C-3 (Table 2.1; Figure 2.4B and 2.8C), shows that processivity is not simply a function of total DNA length. After the first methylation, the enzyme is captured by a distal non-specific segment of flanking DNA mediated by DNA looping (see discussion), as opposed to dissociating from the DNA and

returning to the bulk solution resulting in non-processive catalysis. Thus intersegmental hopping can include transitions between regions on the DNA lacking specific sites, which is likely to be important for proteins searching for scarce sites. However, the enhancement in processivity is most dramatic with site-specific looping (Figure 2.4B).

2.3.6 Dependence of processivity on salt concentration

How variations in salt concentrations impact processivity can reveal mechanistic information about the dynamic association between the protein and the DNA during translocation, as salt screens these interactions (7). Decreases in processivity correlated with increases in salt are explained as a consequence of the protein detaching from the DNA as it moves between sites, and are often ascribed to hopping (12). The anticipated reduction in processivity observed with increasing salt concentrations (Figure 2.9) is consistent with Dam relying on aspects of the hopping mechanism, but reveals an interesting trend: the processivity decreases are similar regardless of the distance between the two sites. Because substrates with longer site to site distances are anticipated to involve more hopping steps, a hopping mechanism predicts these substrates should have more severe decreases in processivity. In contrast, our trend is consistent with Dam undergoing similar paths while moving directly between two regions of the DNA, mediated by DNA looping, in spite of the changes in inter-site distances. We note that the trend continues at low salt concentrations (<30mM), suggesting that Dam has a minor dependency on sliding under these conditions (data not shown).

Figure 2.9. Dam's processivity decreases in the presence of salt. The sensitivity of processivity to salt is evidence of 3-D translocation (black, replot from Figure 2.4B). Importantly, similar salt-dependent decreases in processivity over this range of intersite distances are consistent with intersegmental hopping. Mean and S.D. of 3 trials are plotted.



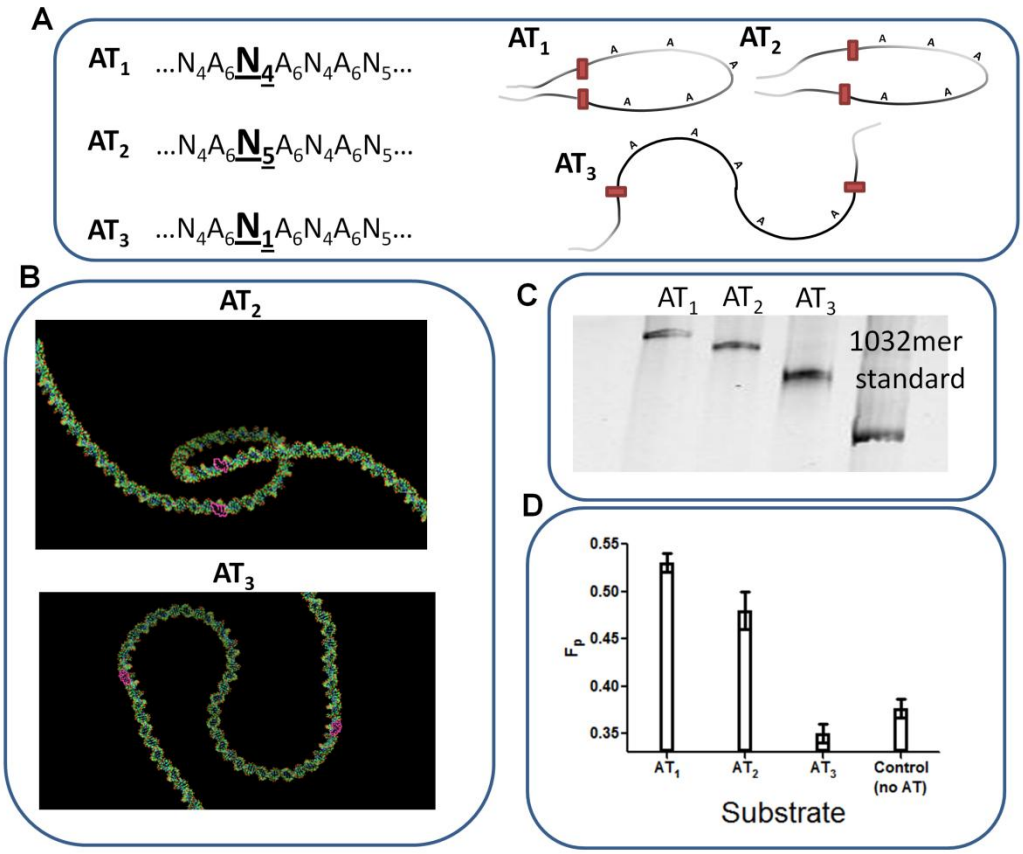
2.3.7 DNA Looping by A-tract architectures

A direct challenge of the intersegmental hopping mechanism, as well as the sliding/hopping mechanisms, relies on the proximal and distal positioning of DNA segments containing Dam sites. We incorporated A-tracts, which have 4-6 adenines in a row that cause distinct bending (~18°) and rigidity to DNA (43), into our substrates. Interestingly, evidence for A-tract mediated bending for DNA was originally observed in its biological context, where A-tracts are involved in kinetoplast DNA (43). The slow migration (by PAGE) of these DNAs was the initial observation that formed the basis for these studies. The bending of the DNA by the A-tracts makes looping around, for example, histones more favorable (see chapter 4). This supports the DNA looping around the histone in spite of the ~150 bp loop distance around the histone (see chapter 4), as in the absence of the A-tracts, such loop formation is highly unlikely and energetically unstable. Several subsequent studies sought to determine the bending properties of these A-tracts, were the amount of adenines and the distance between the centers of the A-tract stretch were varied.

Interestingly, the bending of the DNA by A-tracts can be combined with other protein mediated DNA bends, creating a combinatorial effect to enhance looping. This was demonstrated with the catabolite activator protein (CAP), whose DNA bending properties have been shown to be essential for transcription (46).

Separating each 6 adenine A-tract by intervening 4 and 5 bp stretches of C and G nucleotides achieves an average distance of 10.5 bp between the centers of each bend (Figure 2.10A). When using this phasing, several A-tracts create large global bends within DNA, making distinct macromolecular shapes such as small circles (15 A-tracts, ~150 base pairs long) (44). Well established empirical auto-ligation assays and electrophoretic mobility assays provide the sequence basis for the optimal overall size and A-tract repeat pattern to generate such a circle. This DNA sequence was placed between two GATC sites with the intention of positioning the sites nearly overlapping in space. The presence of flanking DNA separates the sites on different segments of DNA replicating a loop (Figure 2.10A,B).

Figure 2.10. Super-structured substrates affect processivity, affirm intersegmental hopping. **A**) Sequences (AT_{1-3}) used for super-structured substrates. AT_2 incorporates a total of 15 A-tracts between the two GATC sites; the A-tract preceding the bold N is the 7th A-tract for all three substrates. AT_1 and AT_3 incorporate changes to the phasing between the two central A-tracts, dramatically affecting the substrate's structure. The total size of substrates AT_{1-3} is 404-408 base pairs. Schematic representations of the three substrates are shown to the right. AT_1 most closely incorporates a planar circle between the two GATC sites, positioning the sites close together in space. **B**) Modeling of substrates AT_2 and AT_3 confirm anticipated changes in super-structure due to changes in phasing, with the Dam sites in purple (48). **C**) Differential electrophoretic mobility of the three substrates corresponds to the distance between the two GATC sites in space. Included is a 1032 bp DNA standard (no A-tracts) to show the dramatic shifting due to the A-tracts. **D**) Processivity data for the substrates. The closer the sites are in space, the higher the processivity. Included is a DNA control (also no A-tracts) with 174 base pairs between sites and a total size of 408, identical to AT_{1-3}. Mean and S.D. of 3 trials are plotted.



Changing the phasing between two of the central A-tracts, by simply removing 4 bases, dramatically alters the shape of the DNA, making two large bends in opposite directions (Figure 2.10A,B) (45,46). We confirmed the expected conformational changes by electrophoretic analysis. The more planar the circle, the slower it migrates; the more superhelical, the faster it migrates (47) (Figure 2.10C). When the two sites are closest together the mobility is slowest, and when the two sites are farthest apart, and the substrate is in an "s" shape, the mobility is fastest. Interestingly, removing a single bp from the sequence pattern predicted to generate the most planar substrate (43) produced a structure that seems to be more planar (AT_1). DNA modeling also confirms these macromolecular shapes (bend.it DNA modeling program) (48) (Figure 2.10B); however, electrophoresis seems to remain the most reliable characterization technique to differentiate degrees of superhelicity, and in turn the distance between the two GATC sites (Figure 2.10C). As anticipated, the enzyme is most processive when the sites are spatially proximal, suggesting that the enzyme does not track along the DNA to most efficiently go from site to site (Figure 2.10D). Furthermore, when the sites are anchored far away from each other the enzyme is even less likely to travel from site to site than what is seen with a B-form DNA (no A-tract) control that is the same overall size and has the same number of base pairs between the two sites as $AT_{1\text{-}3}$, but includes no A-tracts. Although frequently suggested as a potential translocation mechanism (below), this alternative demonstration that Dam relies on transfers between distal sites on the same DNA provides further evidence for this process.

2.4 Discussion

It is remarkable that proteins can find their sites on DNA within biological time scales in spite of DNA's homogeneity and the presence of other proteins that clutter the DNA (such

as histones and DNA binding/modifying proteins). Identifying the mechanisms involved in this facilitated diffusion process has been a multi-decade undertaking, receiving a recent resurgence of interest. Many translocation studies have elucidated search processes on relatively short DNA segments, but an extrapolation of such mechanisms to large translocations (>1kb) falls short. For example, repair enzymes such as UNG remove bases typically separated by ~10^6 base pairs, and are present at high cellular concentrations (~10^5 copies) (49). Thus, UNG and similar proteins most likely interact with short stretches of DNA (<100 base pairs) by sliding/hopping, thereby obviating the need for a highly processive, *in vivo* mechanism, which has not been demonstrated. Other DNA binding proteins, like Dam, are present in the cell in much smaller quantities (50,51) and therefore each protein is responsible for scanning much larger regions of the DNA. To accomplish this, these proteins must employ larger movements along the DNA. Furthermore, the tendency of DNA to form loops in the cell must be considered for *in vivo* protein site finding and processive catalysis (9). Several reports concerning cellular DNA conformation, protein roadblocks, and sliding/hopping kinetics conclude that facilitated diffusion involves some combination of sliding/hopping and ambiguously described 3-D movements, which intersegmental hopping accommodates (3,8,9,33). Taken together, intersegmental hopping represents a plausible, intuative strategy for efficient long range site location. The following comments on previous translocation studies support this notion.

Recent experiments suggest that EcoRI ENase (37), EcoRV ENase (38,39), and AAG (40) can utilize intersegmental hopping. For example, by mechanically controlling DNA coiling, van den Broek *et. al* show (35) that passive DNA looping enhances the site location kinetics of EcoRV ENase, allowing it to transfer between distal non-specific regions of the

DNA, similar to our result in Figure 2.8C. Interestingly, the heavy reliance on sliding/hopping by EcoRV ENase obscures intersegmental hopping when probed using processivity assays (12). Therefore, the extent that these enzymes actually utilize this mechanism for translocation remains unclear, and while these reports are very helpful in providing precedence for intersegmental hopping, the contribution of this mechanism to processivity has not been previously demonstrated.

We emphasize a clear distinction in showing whether a protein can utilize a certain mechanism, and whether a protein will utilize that mechanism in spite of the availability of other mechanisms (our results here correspond to the latter case, and we believe we are this first to demonstrate this for the intersegmental hopping mechanism using processivity). For example, if EcoRI ENase or EcoRV ENase utilize intersegmental hopping much more so that sliding and/or hopping mechanisms, then this will be revealed in its processivity plot (Figure 2.4B). Since this is not the case, EcoRI ENase and EcoRV ENase likely do not use this DNA looping mediated translocation mechanism to the extent of Dam. We highlight that since Dam is shown to use intersegmental hopping to enhance processivity, that this is a novel and unique result.

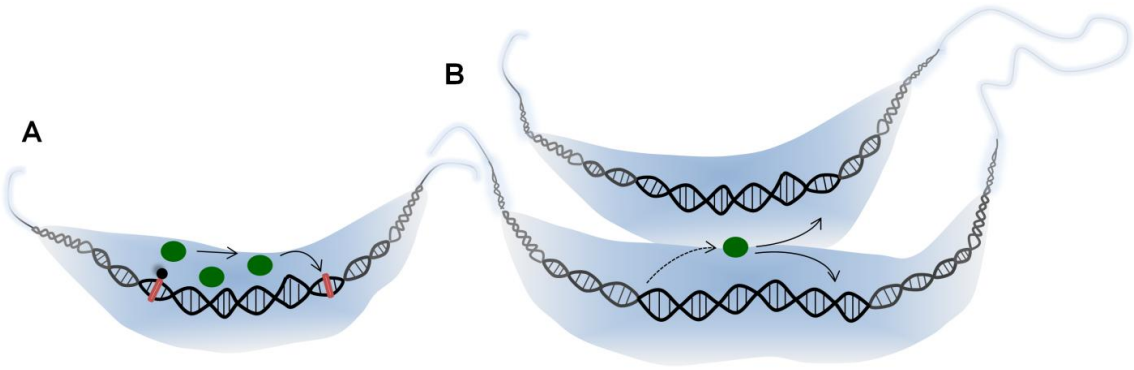
Other reports cannot be reconciled by sliding/hopping models, and are suggestive of intersegmental hopping (7,52). For example, BbvCI ENase is roughly equally processive on DNA (301bp) with distances between the two sites ranging from 30 to 75 bp (52). One explanation for the unchanging processivity is the contribution from non-specific looping, which would be predicted to affect processivity given the total size of the DNA. Furthermore, salt has a similar impact on BbvCI ENase processivity trends as Dam (Figure 2.9): when the salt level is increased, the processivity is lowered equally for the series of

substrates with 30-75bp intersite spacings (52). For example, under 150mM NaCl, processivity for a substrate with 75 bp between sites is only slightly less than when 30 bp separates the sites. But processivity drops dramatically between 0 and 150mM NaCl. We stress that this data is inconsistent with the conclusion of their study, which invokes that BbvCI utilizes a combination of hopping and sliding, with hopping dominating with the larger intersite spacings. Both the sliding and hopping models predict that processivity should decrease with increases in the intersite distances. Particularly for the high salt case, since processivity doesn't lower when the intersite distances increase, reconciliation with a sliding or hopping mechanism seems impossible. This oversight was similarly made by O'Brien and colleagues, who were looking at processive catalysis of AAG (7): processivity for two substrates, one with an intersite spacing of 25 bp and another with an intersite spacing of 50bp, remain equal under a range of NaCl concentrations of 0-300mM. Again, one would expect that under high salt, given the hopping model, that processivity would be lowered when the intersite spacing increases. Interestingly, AAG was later reconciled to indeed utilize a mechanism we would consider to be very similar to that of intersegmental hopping. Finally, the observations that non-bridging ENases AscI and Tth111I (53,54) show significant processivity with extremely long site to site distances are also challenging to account for with a sliding/hopping mechanism. Although this data is a part of a larger study involving enzymes that oligomerize by DNA looping to achieve processive catalysis, we stress that the authors state that ENases AscI and Tth111I act as monomers similarly to Dam.

What remains unclear is how Dam (and other proteins) carries out these translocations. Proteins can meander within a region spanning several nanometers away from DNA as they

move during hopping (7,8); for example AAG can hop over a non catalytic EcoRI ENase roadblock, which only partially perturbs its ability to undergo processive catalysis (7). Dam's processivity is unaffected by a similar roadblock (see chapter 4), which is further evidence in favor of intersegmental hopping. Also, several proteins (including Dam) can switch between DNA strands within a single duplex (16,30,55,56). For this, similar to hopping, a transient protein intermediate exists unattached to the DNA but poised for reassociation to either strand. Many proteins clearly display interesting acrobatics surrounding the DNA, generally considered hopping, the details of which remain unclear. Perhaps intersegmental hopping involves the "capture" of the translocating protein by looping DNA, during these short excursions away from the DNA in which binding to the looped DNA becomes as probable as rebinding the original region (Figure 2.11). Notably, these transitions involve countless intermediates (3). Site specific intersegmental hopping likely involves some short sliding/hopping around the sites during the transitions. Given that intersegmental hopping relies on short excursions away from the DNA, it can be considered an intuitive yet mechanistically powerful extension of hopping. Intersegmental hopping is also similar to "jumping" or 3-D diffusion, but these have been described to include large movements throughout space not associated with any DNAs, which we feel is not entirely appropriate for the mechanism described here. What seems unlikely is that the proteins discussed here exploit a secondary DNA binding site, and are therefore incapable of intersegmental transfer.

Figure 2.11. Intersegmental hopping mechanism. **A**) Proteins (green circles) can use short range sliding or hopping to follow along the trace of the helix between sites, staying within the (blue) region where re-association remains likely. **B**) When DNA looping allows two helices to become proximal, the overlapping of these regions can facilitate intersegmental hopping from one region of the DNA to another.



Our results suggest that Dam manifests two forms of intersegmental hopping. The first (site-specific) is manifested when intersegmental hopping and subsequent short sliding/hopping ends with the protein on its site, and the second (non-specific) involves intersegmental hopping between nonspecific segments that doesn't immediately result in specific binding; the former allows further enhancement of processivity for Dam, peaking at site separations of approximately 500-600 bp. Interestingly, the maximum processivity values from Figure 2.8C (~0.5) match well with what appears to be a plateau in Figure 2.4B, further suggesting both types of intersegmental hopping mechanisms (specific and non-specific) contribute to our processivity trends. Importantly, the effects of non-specific intersegmental hopping are not predicted to decrease as DNA length increases, and therefore this mechanism is able to be a source of movement spanning several kb, as has been shown previously (38).

By construction, k' in Equation 12 sequesters site-specific intersegmental hopping from other mechanisms contributing to processivity. This has allowed us to model the effect of

varying site separation in an elementary way and to confirm plausibility of the specific intersegmental hopping mechanism without fully elucidating all mechanisms that conspire to yield the measured values of F_p. The price paid for the simplicity of the model is that it cannot quantitatively predict the relative importance of other translocation mechanisms we know must also contribute. However, non-specific intersegmental hopping certainly appears to contribute significantly to the magnitude of α. The inability of Dam to significantly utilize sliding/hopping mechanisms is further highlighted by considering substrate 5C-1 (Figure 2.8C), which is too small for either specific or non-specific intersegmental hopping, and is therefore nearly unprocessive; this suggests that sliding/hopping likely don't significantly contribute to the magnitude of α. A translocation model that relies to various extents on sliding, hopping and both types of intersegmental hopping has yet to be developed, and could presumably help understand both our data and previous data that is poorly fit to current models (7,52-54). We anticipate that the relative contributions from different translocation mechanisms will be protein specific.

Biological context ultimately drives the combination of translocation mechanisms displayed by a particular protein. What seems to be unique to Dam is that it inefficiently translocates (and methylates) between closely spaced sites. This property reveals Dam's reliance on intersegmental hopping and is consistent with its proposed mechanism for gene regulation (19). The maintenance of the epigenetic state of the well-studied *pap* operon involves the differential methylation of its two closely spaced GATC sites (57).This supports a non-processive mechanism, where only one of the two GATC sites is methylated following replication. The unique sequence flanking each GATC sites leads to decreasing processivity (19), and our data here shows the clustering of the GATC sites is also

important. Additionally, the clustering of GATC sites supports intrasite methylation (16) (the methylation of both adenines within a single palindromic GATC site during one binding event) which possibly distracts from processive methylation of closely spaced sites. Dam clearly has evolved an elegant mechanism to satisfy its diverse cellular roles.

Instead of short, rigorous yet counterproductive searches on small regions of DNA by sliding/hopping, Dam likely relies on a looping process to translocate and processively methylate sites which are separated by hundreds of bps for its role in mismatch repair. While the methylation of every GATC site in the *Escherichia coli* genome is not critical for this function (the mismatch repair pathway is sufficient when methylated sites are separated by 1 kb (58,59)), DNA repair enzymes – for example – must search local regions of the DNA more meticulously as locating every target is important. Differential consequences of skipping over sites demonstrate again the necessity for diverse translocation strategies. High levels of repair proteins accommodate this. While intersegmental hopping is highly efficient for large translocations, the extent to which other proteins rely on this has not been demonstrated or broadly accepted as the appropriate model systems have not yet been explored. However, prior reports demonstrate that translocation mechanisms other than intersegmental hopping contribute to site finding, and we highlight that many aspects of Dam appear to be unique. Here we show a novel relationship between processivity and DNA confirmation. Intersegmental hopping is potentially important for low copy proteins faced with locating relatively rare sequences, such as many transcription factors (9).

2.5 Materials and Methods

Processivity assay: (Dam) All reactions were done in MRB (methylation reaction buffer:100mM Tris H-Cl, pH8.0, 10 mM EDTA, 1mM DTT, 0.2mg/mL BSA(bovine serum

albumin)) with 400nM DNA, 7nM Dam, 0.2mg/mL BSA, and 30 µM AdoMet (S-adenosyl methionine, SAM). Dam was diluted in Protein Dilution Buffer (20mM potassium phosphate, 200mM NaCl, 0.2mM EDTA (ethylenediaminetetraactetic acid) 0.2 mg/mL BSA 2mM DTT (dithiothreitol) and 10% glycerol). Aliquots of the reaction mixture were heat inactivated in 75 degree water, which has previously been shown to be effective (16). After slow cooling, NEB DpnII Buffer and DpnII enzyme was added and the mixture was incubated at 37 degrees for 2 hours. The reaction products were analyzed using PAGE (20%-5% (depending on substrate length) 29:1 acrylamide: bisacrylamide) Processivity assays from Figure 2.10 were done with the addition of 15 mM MgCl (the presence of EDTA in the reaction this gives an effective concentration of 5 mM MgCl) at room temperature to accommodate optimal A-tract bending conditions. Sybr Au was purchased from Invitrogen and was used as instructed.

(EcoRI ENase) All reactions were done in ERB (endonucleases reaction buffer: 20 mM Tris pH 7.4, 5 mM $MgCl_2$, 0.2 mM EDTA, .05 mg/ml BSA) with 400nM DNA, 65mM NaCl, and 26 nM EcoRI ENase at 37 degrees. Reactions were quenched in 50mM EDTA then run on a PAGE gel as above.

Enzyme expression and purification: Dam was purified exactly as before (16). EcoRI ENase was purchased from NEB.

Tritium Assay: All steady-state reactions were done in MRB with 400nM DNA, 1.75-14nM Dam, 0.2mg/mL BSA, and 30 µM AdoMet (6Ci/mmol mixture of unlabeled and [^{3}H]methyl labeled, Perkin Elmer) at 37 degrees centigrade, and were quenched with1% SDS (sodium dodecyl sulfate) then spotted on DE81 filter paper. The paper was washed three times with a 50mM KH_2PO_4 buffer, once in 80% ethanol, once in 95% ethanol, and

was dried in diethyl-ether; all for five minutes. Papers were dried and submerged in BiosafeII scintillation fluid. Tritium levels were quantified using a Beckman Coultier LS6500 scintillation counter and converted to nM DNA product.

Data analysis: The processivity values (Table 2.1) were derived from a least squares fitting using Microsoft Excel. The modeling (Figure 2.4B, Figure 2.5) was executed using Wolfram Mathamatica 9. Gels were scanned on a Typhoon Phosphoimager (GE)

DNA substrates: All restriction endonucleases for cloning were purchased from NEB. All synthetic DNA substrates and primers were purchased from IDT and were re-suspended in TE buffer (10mM Tris, pH 7.5, 1mM EDTA). When necessary, DNA was annealed with its reverse complement in a 1:1 mixture for 5 minutes at 95 degrees, and allowed to slow cool to room temperature (~5 hours). Annealing was verified by PAGE. The production of several PCR amplicons used the following strategy: A synthetic oligonucleotide with two GATC sites (below), and two restriction sites between the GATC sites was cloned into plasmid pBR322 (NEB), making a new plasmid, pBR322GATC. The same strategy was used to generate the EcoRI Enase substrates making another plasmid, pBR322GAATTC. Several spacers were generated by PCR (from a different region of the pBR322 plasmid) and were cloned into pBR322GATC, generating several different plasmids with different distances between the GATC sites. Amplicons were generated by PCR using a single set of primers for substrates used in Figure 2.4B, Figure 2.8C). pBR322GATC and pBR322GAATTC were PCR amplified with different primers to adjust the spacings from the GATC sites (and GAATTC sites) to the ends of the DNA (data from Figure 2.8C).

The following substrate (described above) was annealed and cloned into the plasmid pBR322 at the EcoRI and HindIII sites, 5’-

aattcggtgatctcgtcga*cccggg*agctggtagtatgc*ccatgg*ttcgatcggatgcca-3’; 5’-agcttggcatccgatcgaa*ccatgg*gcatactaccagct*cccggg*tcgacgagatcaccg-3’, making, pBR322GATC. The following substrate was annealed then cloned into the plasmid pBR322 also at the EcoRI and HindIII sites, 5’-agcttggcattgaattctaccatgggcatactaccagctcccgggtcggctcgatcaccg-3’ ; 5’-aattcggtgatcgagccgacccgggagctggtagtatgcccatggtagaattcaatgcca-3’, making pBR322GAATTC. The cloned, synthetic insert (pBR322GATC) had additional cloning sites within it: XmaI and NcoI (italicized). These sites were used to insert PCR purified spacers between the two GATC site(s) (underlined). Upon PCR amplification, the spacers were digested with XmaI and NcoI to generate overhangs. The spacers were generated by PCR from the plasmid pBR322 with restriction sites using the same forward primer: 5’- att *cccggg* ggctaccctgtggaacacct - 3’, with different reverse primers for each sized spacer, starting with substrate 3B-2:

Substrates: name, size between sites, total size

4B-1, 36, 270

(no spacer)

4B-2, 134, 364:

5’- taatccatggtactggaacgttgtgagggt-3’

4B-3, 194, 428:

5’- taatccatggtaccgatgaaacgagagagg-3’

4B-4, 234, 468:

5’- taatccatgggcctccgtgtaaaggggatt-3’

4B-5, 284, 518:

5’- taatccatggtgataaagcgggccatgtta-3’

4B-7, 395, 629:

5’- taatccatgggcagctgcggtaaagctcat-3’

4B-8, 482, 716:

5’- taatccatggcccggcatccgcttacagac-3’

4B-9, 605, 839:

5’- taatccatggcaatctgctctgatgccgca-3’

4B-10, 798, 1032:

5’-taatccatggcatgttctttcctgcgttatcccc-3’

As a control, a spacer was constructed from a different region of the plasmid:

Forward: 5’-attcccgggatagttgcctgactccccgt-3’

4B-6, 335,569:

5’- taatccatggcgttgggaaccggagctgaa-3’

Amplicons used in Figure 2.10 were constructed using the same strategy, except the spacers were made using three ligated synthetic oligonucleotides (“1-3”). The only differences between the substrates come from oligonucleotide “2”. The numbers below refers to the order the inserts were annealed in starting with the 5’ direction. “rev” refers to the reverse compliment of the oligonucleotide.

AT_2

1) 5’-ccgggaaaaaaccggcaaaaaacgcgaaaaaagcccgaaaaaacccgaaaaaacgccg-3’

2) 5’-aaaaaagcgcaaaaaacgggcaaaaaacgccaaaaaaggccgaaaaaaccgga-3’

3) 5’-aaaaaccgggaaaaaacgggaaaaaagcccgaaaaaagccgaaaaaac-3’

1rev) 5’- ttttttcgggtttttcgggctttttcgcgtttttgccggtttttc-3’

2rev) 5’- ttttttcggccttttttggcgtttttgcccgtttttgcgctttttcggcg-3’

3rev) 5’- catggtttttcggctttttcgggctttttcccgtttttcccggtttttccgg-3’

AT_1: same 1 and 3 from AT_2

2_1) 5’-aaaaaagcgcaaaaaacggcaaaaaacgccaaaaaaggccgaaaaaaccgga-3’

$2rev_1$) 5’-ttttttcggccttttttggcgtttttgccgtttttgcgctttttcggcg-3’

AT_3: same 1 and 3 from AT_2

2_3) 5’-aaaaaagcgcaaaaaacaaaaaacgccaaaaaaggccgaaaaaaccgga-3’

$2rev_3$) 5’- ttttttcggccttttttggcgtttttgtttttgcgctttttcggcg-3’

Amplicons with 115/119 base pairs flanking GATC sites were amplified from each plasmid using primers: (forward) 5’- gggttccgcgcacatttccc-3’ and (reverse) 5’-(Fl) ccagggtgacggtgccgagg-3’. The previous primers were exclusively used for the data in Figure 2.4B and 2.10. Substrate 5C-1 is 5’ (Fl)-atcattgcgatcttggacttggaacgaattcgatatcggttccaatgcagatcttggcca -3’annealed to its reverse compliment. The following list of primers was used in Figure 2.8C. Amplicons with 30 base pairs flanking GATC sites (substrate 5C-2) were amplified from each plasmid using primers: (forward) 5’- ggcgtatcacgaggccctttcg -3’, and (reverse) 5’-(Fl) cgttagcaatttaactgtgataaactaccgc -3’. Amplicons with 300 base pairs flanking GATC sites (substrate 5C-3) were amplified from each plasmid using primers: (forward) 5’- gcatcttt tactttcacc agcg-3’, and (reverse) 5’- (Fl)ggctccaagtagcgaagcgagc-3’. PCR amplicons were purified using the Agilent PCR clean-up kit and ethanol precipitated to achieve desired concentrations.

2.6 References

1. Berg, O. G., Winter, R. B. & von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry.* **20**, 6929-6948

2. Halford, S. E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* **37**, 343-348

3. Hammar, P. *et al*. (2012) The *lac* repressor displays facilitated diffusion in living cells. *Science.* **336**, 1595-1598

4. Iwahara, J. & Clore, G.M. (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*. **440**, 1227-1230

5. Esadze, A. & Iwahara, J. (2014) Stopped-flow fluorescence kinetic study of protein sliding and intersegmental transfer in the target DNA search process. *J. Mol. Biol.* **426**, 230-244

6. Zhou, H. X. (2011) Rapid search for specific sites on DNA through conformational switch of non-specifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8651-8656

7. Hedglin, M. & O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. *ACS Chem. Biol.* **5**, 427-436

8. Schonhoft, J. D. & Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA. *Nat. Chem. Biol.* **8**, 205-210

9. Li, G. W., Berg, O. G. & Elf, J. (2009) Effect of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **5**, 294-297

10. Wion, D. & Casadesus, J. (2006) N^6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* **4**, 183-192

11. Heithoff, D. M., Sinsheimer, R. L., Low, D. A. & Mahan, M. J. (1999) An essential role for DNA adenine methylation in bacterial virulence. *Science*. **284**, 967-970

12. Stanford, N. P., Szczelkun, M. D., Marko, J. F. & Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* **19**, 6546-6557

13. Porecha, R. H. & Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10791-10796

14. Halford, S. E. & Marko J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res*. **32**, 3040-3052

15. Hu, T., Grosberg, A. Y. & Shklovskii, B. I. (2006) How proteins search for their specific sites on DNA: The role of DNA confirmation. *Biophys. J.* **90,** 2731-2744

16. Pollak, A. J. & Reich, N. O. (2012) Proximal recognition sites facilitate intrasite hopping by DNA adenine methyltransferase: mechanistic exploration of epigenetic gene regulation. *J. Biol. Chem.* **287**, 22873-22881

17. Fersht, A. (1998) *Structure and Mechanism in Protein Science*, W. H. Freeman, New York.

18. Langowski, J., Alves, J., Pingoud, A. & Maass, G. (1983) Does the specific recognition of DNA by the restriction endonuclease EcoRI involve a linear diffusion step? Investigation of the processivity of the EcoRI endonuclease. *Nucleic Acids Res.* **11**, 501-513

19. Peterson, S. N. & Reich, N. O. (2006) GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence *pap* expression. *J. Mol. Biol.* **355**, 459-472

20. Ryabinina, O. P., Minko, I. G., Lasarev, M. R., McCullough, A. K. & Lloyd, R. S. (2011) Modulation of the processive abasicsite lyase activity of a pyrimidine dimer glycosylase. *DNA Repair* **10,** 1014-1022

21. Shore, D., Langowski, J. & Baldwin, R. L. (1981) DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4833-4837

22. Ringrose, L. & Chabanis, S. (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO J.* **18**, 6630-6641

23. Shimada, J. & Yamakawa, H. (1984) Ring closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules.* **17**, 689-698.

24. Hagerman, P. J. & Ramadevi, V. A. (1990) Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. *J. Mol. Biol.* **212**, 351-362

25. Geggier, S. & Vologodskii, A. (2010) Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15421-15426

26. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. (1997) Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6185-6190

27. Wentzell, L. M. & Halford, S. E. (1998) DNA looping by the Sfi I restriction endonuclease. *J. Mol. Biol.* **281**, 433-444

28. Catto, L. E., Ganguly, S., Milsom, S. E., Welsh, A. J. & Halford, S. E. (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res.* **34**, 1711-1720

29. Friedhoff, P., Lurz, R., Luder, G. & Pingoud, A. (2001) Sau3AI, a monomeric type II restriction endonuclease that dimerizes on the DNA and thereby induces DNA loops. *J. Biol. Chem.* **276**, 23581-23588

30. Coffin, S. R. & Reich, N. O. (2009) Escherichia coli DNA adenine methyltransferase: intrasite processivity and substrate-induced dimerization and activation. *Biochemistry* **48**, 7399-7410

31. Horton, J. R., Liebert, K., Bekes, M., Jeltsch, A. & Cheng, X. (2006) Structure and substrate recognition of Escherichia coli DNA adenine methyltransferase. *J. Mol. Biol.* **358**, 559-570

32. Vuzman, D., Azia, A. & Levy, Y. (2010) Searching DNA via a "monkey bar" mechanism: The significance of disordered tails. *J. Mol. Biol.* **396,** 674-684

33. Iwahara, J., Zweckstetter, M. & Clore, G. M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15062-15067

34. Doucleft, M. & Clore, G. M. (2008) Global jumping and domain-specific intersegmental transfer between DNA cognate site of the same multidomain transcription factore Oct-1. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13871-13876

35. Iwahara, J. & Clore, G.M. (2006) Direct observations of enhanced translocations of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J. Am. Chem. Soc.* **128**, 404-405

36. Tjong, H. & Zhou, H. X. (2007) DISPLAR: An accurate method for predicting DNA-Binding sites on protein surface. *Nucleic Acids Res.* **35**, 1465-1477

37. Sidorova, N. Y., Scott, T. & Rau, D. C. (2013) DNA concentration dependent dissociation of EcoRI: direct transfer or reaction during hopping. *Biophys. J.* **104**, 1296-13039

38. van den Broek, B., Lomholt, M. A., Kalisch, S. M. J., Metzler, R. & Wuite, G. J. L. (2008) How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1578-15742

39. Gowers, D. M. & Halford, S. E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO. J.* **22**, 1410-1418

40. Hedglin, M., Zhang, Y. & O'Brien, P. J. (2013) Isololating contributions from intersegmental transfer to DNA searching by Alkyladenine DNA Glycosylase. *J. Biol. Chem.* **288**, 24550-24559

41. Rau, D. C. & Sidorova, N. Y. (2010) Diffusion of the restriction nuclease EcoRI along DNA. *J. Mol. Biol.* **395**, 408-416

42. Terry, B. J., Jack, W. E. & Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonucleases. Non-specific interactions in EcoRI catalysis. *J. Biol. Chem.* **260**, 13130-13137

43. Koo, H. S., Wu, H. M., Crothers, D. M. (1986) DNA bending at adenine thymine tracts. *Nature* **320**, 501-506

44. Koo, H. S., Drak, J., Rice, J. A. & Crothers, D. M. (1990) Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry* **29**, 4227-4234

45. Zinkel, S. S. & Crothers, D. M. (1987) DNA bend direction by phase sensitive detection. *Nature* **328**, 178-181

46. Kahn, J. D. & Crothers, D. M. (1992) Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6343-6347

47. Stefl, R., Wu, H., Ravindranathan, S., Sklenar, V. & Feigon, J. (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1177-1182

48. Vlahovicek, K., Kajan, L. & Pongor, S. (2003) DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* **31**, 3686-3687

49. Friedman, J. I. & Stivers, J. T. (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957-4967

50. Ghaemmaghami, S. *et al.* (2003) Global analysis of protein expression in yeast. *Nature* **425**, 737-741

51. Elf, J., Li, G. W., Xie, X. S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell, *Science* **316**, 1191-1194

52. Gowers, D. M., Wilson, G. G. & Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15883-15888

53. Bilcock, D. T., Daniels, L. E., Bath, A. J. & Halford, S. E. (1999) Reactions of type II restriction endonucleases with 8-base pair recognition sites. *J. Biol. Chem.* **274**, 36379-36386

54. Gormley, N. A., Bath, A. J. & Halford, S. E. (2000) Reactions of BglI and other type II restriction endonucleases with discontinuous recognition sites. *J. Biol. Chem.* **275**, 6928-6936

55. Zaremba, V. *et al*. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′–5′ and 5′–3′ strands by rotating a single active site. *Nucleic Acids Res.* **38,** 2399-2410

56. Sasnauskas, G., Kostiuk, G., Tamulaitis, G. & Siksynys, V. (2011) Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.* **39**, 8844-8856

57. Hernday, A. D., Braaten, B. A. & Low, D. A. (2003) The mechanism by which DNA adenine methylase and papI activate the pap epigenetic switch. *Mol. Cell* **12,** 947-957

58. Pluciennik, A. & Modrich, P. (2007) Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12709-12713

59. Kunkel, T. A. & Erie, D. A. (2005) DNA Mismatch Repair. *Annu. Rev. Biochem.* **74,** 681-71

III. Distinct facilitated diffusion mechanisms by *E. coli* Type II restriction endonucleases

3.1 Abstract

The passive search by proteins for particular DNA sequences involving non-specific DNA is essential for gene regulation, DNA repair, phage defense, and diverse epigenetic processes. Distinct mechanisms contribute to these searches and it remains unresolved which mechanism or blend of mechanisms best suits a particular protein and more importantly, its biological role. To address this we compare the translocation properties of two well-studied bacterial restriction endonucleases (ENases), EcoRI and EcoRV. These dimeric, magnesium-dependent enzymes hydrolyze related sites (EcoRI ENase: 5'-GAATTC-3' and EcoRV ENase: 5'-GATATC-3') leaving overhangs and blunt DNA segments respectively. Here we demonstrate that the extensive sliding by EcoRI ENase, involving sliding up to ~600 bp prior to dissociating from the DNA, contrasts with a larger reliance on hopping mechanism(s) by EcoRV ENase. The mechanism displayed by EcoRI ENase results in a highly thorough search of DNA whereas the EcoRV ENase mechanism results in an extended, yet less rigorous interrogation of DNA sequence space. We describe how these mechanistic distinctions are complemented by other aspects of these endonucleases, such as the ten-fold higher *in vivo* concentrations of EcoRI ENase compared to EcoRV ENase. Further, we hypothesize that the highly diverse enzyme arsenal which bacteria employ against foreign DNA involves seemingly similar enzymes which rely on distinct but complementary search mechanisms. Our comparative approach reveals how different proteins utilize distinct site locating strategies.

3.2 Introduction

Bacterial restriction endonucleases are broadly dispersed (1) and provide a robust safeguard against various foreign nucleic acids. While other strategies contribute toward this protection (e.g., CRISPR (2), RecBCD (3), and abortive infection systems (4)), the diversity of endonucleolytic-based strategies speaks to their importance. Many endonucleases (ENases) are coupled with a methyltransferase which together form a type II Restriction Modification (R-M) system, one of four types of R-M systems (1,5). The ENase(s) recognize and cleave unmethylated sites in invading DNA, while the methyltransferase(s) protects those sites on the bacterial genome. A delicate interplay between these two competing enzyme types is necessary for this protection, preventing, for example, the ENase from cutting genomic DNA, yet allowing certain foreign DNA elements to be incorporated into the bacterial genome, providing genetic diversity. Among the superfamily of ENases, Type II's, which typically cut 4-8 base pair (bp) sites, are the most diverse and abundant (6). Here we compare two well-studied Type II ENases, EcoRI ENase (EcoRI) (5'-GAATTC-3') and EcoRV ENase (EcoRV) (5'-GATATC-3'), both of which are Mg^{2+} dependent homodimers found in *E. coli* (6). We are interested in defining the mechanistic distinctions between these ENases, and how any distinctions may help understand why most bacteria contain several seemingly redundant R-M systems (1).

The vast majority of DNA binding/modifying proteins move along DNA in search of their recognition sites without relying on an external energy source, a process called facilitated diffusion (7,8,9). Some proteins are present at a few copies per cell (~100) whereas other proteins are present at very high levels (e.g., > 300,000 per cell) (10,11,12); moreover, some proteins locate sites that occur frequently (e.g., every 20-100 bp) whereas others locate sites separated by millions of bp (12,13). Distinct mechanisms are therefore

likely to contribute to these facilitated diffusion processes (Figure 3.1); however, experimental evidence clearly segregating or directly comparing these mechanisms is limited. Processive catalysis, as measured by the efficiency of modifications involving a second site, provides insight into the underlying translocation mechanism (14,15,16). ENases that cut both sites during one binding event, where the enzyme travels to the second site without returning to the bulk solution following catalysis at the first site, are considered processive. How processivity changes with different site to site (intersite) spacings between two sites is critical for deducing translocation mechanisms and remains a powerful technique to do so (14,15,16). Sliding is one such mechanism, where enzymes linearly diffuse along the contour of DNA (Figure 3.1). The ratio of sliding kinetics (k_{slide}) to dissociation kinetics (k_{off}) scales with the square of the bp (n) between the two sites

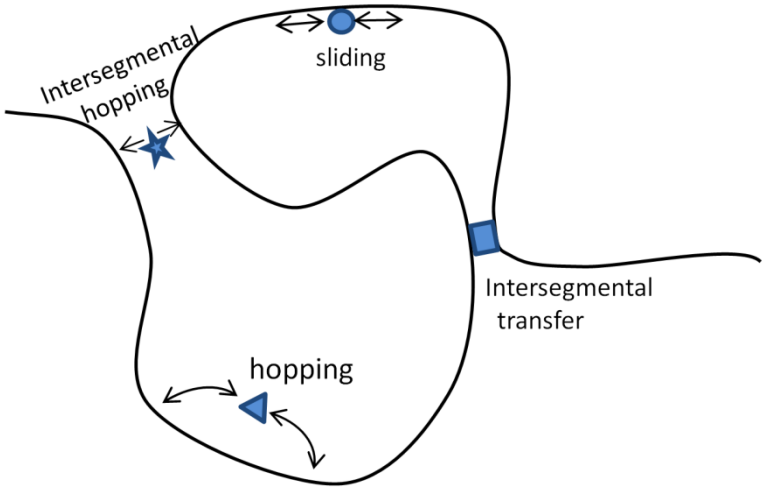
$$F_p = (k_{slide} / (k_{slide} + k_{off}))^{n\wedge 2} \quad (1)$$

making processivity (F_p) at, for example > 250 bp increasingly unlikely (8,14). Hopping involves rapid dissociation and reassociation within a poorly defined region surrounding the DNA (Figure 3.1). Hopping scales as the inverse of r, which is the intersite distance defined by the worm-like chain model, and a, which is the distance the protein slides between each hop (14,16).

$$F_p = (a/r) \quad (2)$$

Processivity is typically less sharply dependent on the intersite spacings for hopping than sliding. As hopping involves larger displacements along the DNA than sliding, it is often proposed to lead to more efficient long distance search mechanism (8). Importantly, studies of proteins sliding and hopping largely rely on intersite spacings of 10 bp to several hundred bp, and *in vivo* search events often involve thousands of bp between sites.

Figure 3.1. Known facilitated diffusion mechanisms for DNA binding/modifying proteins. The circle shows sliding, where the protein makes a series of single bp shifts, usually redundantly sampling small regions of DNA. The triangle represents hopping, where the protein undergoes a series of disassociation and reassociation steps, allowing larger areas of the DNA to be sampled. The star shows intersegmental hopping, where the looping of DNA captures the hopping protein, allowing it to transfer between distal DNA segments. The square shows intersegmental transfer, where the protein transfer involves simultaneous binding of two DNA segments.



Evidence has been recently reported for a translocation mechanism which relies on hopping but includes a distal DNA segment being positioned in proximity to the protein, allowing it to transfer between DNA duplexes when hopping, providing a means to efficiently translocate between DNA segments (17-20). We deemed this "intersegmental hopping (17)," and it is distinct from intersegmental transfer, which requires an intermediate where proteins or protein complexes simultaneously bind two DNA segments (21,22). Intersegmental hopping occurs optimally when the segments within the same strand are separated by 500-600 bp, as shown by our previous work (17). This bp dependence corroborates well with the probability of DNA loop formation of sites (23,24). We suggested that the relative reliance on sliding, hopping and/or intersegmental hopping of a particular protein is driven by its particular biological context, including target site density, cellular

protein (of interest) levels, and the relative consequence of site finding (17). For example, *E. coli* DNA adenine methyltransferase (Dam) primarily utilizes intersegmental hopping as only ~100 Dam enzymes methylate ~20,000 sites along the entire bacterial genome. Since skipping sites is relatively inconsequential for Dam (see discussion), sliding/hopping short distances would be unnecessarily redundant and indeed is not significantly utilized (17). Here we are interested in whether EcoRI and EcoRV, both of which recognize and cut six bp recognition sites, display distinct search mechanisms and if those distinctions can further elucidate details concerning translocation mechanisms, and help understand a coordinated host defense strategy.

Extensive studies of EcoRI and EcoRV reveal unresolved questions about their translocation properties (14,25,26). Modrich *et al*. found that EcoRI is more processive with circularized versus linear DNA (27). These results have been interpreted to involve hopping or "3D translocation" (possibly related to intersegmental hopping) (14), but extensive sliding cannot be dismissed. Rau *et al.* used DNA segments with under 100 bp intersite spacings as a basis for extrapolating a sliding mechanism out to ~400 bp (28). EcoRV utilizes a combination of short range sliding and longer range hopping mechanisms, as deduced by standard processivity experiments and equations (Equations 1 and 2) (14). Although EcoRV can carry out what we call intersegmental hopping (18,19), the relative importance of this translocation mechanism in comparison to sliding and hopping mechanisms remains inconclusive. Interestingly, both EcoRI and EcoRV can translocate hundreds of bp as part of their search strategy (14,27). Here we investigate how they differ in the particular blend of the underlying site finding mechanisms. Importantly, both the

extent and the comparative nature of our experiments provide several new insights to these well-studied systems.

3.3 Results

3.3.1 Processivity Assay

Our processivity values for EcoRI are derived using a multiple turnover steady-state assay with 400 nM DNA and 1 nM EcoRI enzyme (dimers) to minimize situations where two enzymes are bound to a single DNA molecule, as was shown previously in Pollak *et al.* (17). Each two site substrate is symmetric with respect to the amount of DNA flanking each site to avoid kinetic site preferences. The DNA fragments are separated and quantified with by PAGE. The fluorescein label is on only one end of the DNA to simplify the analysis. Processivity values are based on the relative amounts of the top band (Uc), the middle band (Sc'), and the bottom band (Dc*) at each time point (Figure 3.2), which provide the relative amounts of uncut, singly cut and doubly cut products upon subsequent manipulation (Eq. 3-5). The presence of the unlabeled "ghost bands" is accounted for by multiplying Sc' by 2, revealing the amount of singly cut products. Dc* is subtracted by Sc' to account for the presence of by-product of the singly cut ghost band, revealing the amount of doubly cut product (17).

Total density per lane: $T = Uc+2Sc'+Dc^*-Sc$ (3)

$$T = Uc+Sc'+Dc^* \quad (3a)$$

$$\text{Singly cut (Sc)} = 2Sc'/T \quad (4)$$

$$\text{Doubly cut (Dc)} = (Dc^*-Sc')/T \quad (5)$$

These values are fit to a sequential reaction mechanism, revealing k_1, (the first order rate constant from uncut to singly cut) k_2, the first order rate constant from singly cut to doubly

cut (Equations 6-7) (30), and finally F_p (Equation 8) (17).

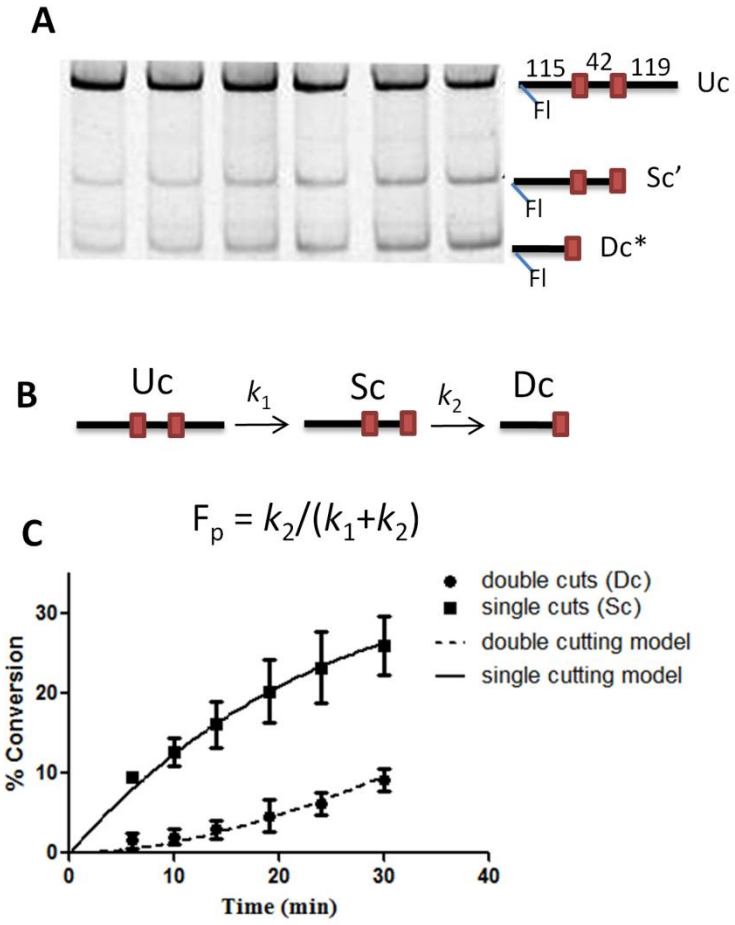
$$\text{Singly cut (S)} = \frac{k_1}{(k_2-k_1)} \times (e^{-tk_1} - e^{-tk_2}) \quad (6)$$

$$\text{Doubly cut (D)} = 1 + \frac{1}{(k_1-k_2)} \times (k_2 e^{-tk_1} - k_1 e^{-tk_2}) \quad (7)$$

$$F_p = \frac{k_2}{k_1+k_2} \quad (8)$$

For unprocessive catalysis each cutting event is uncorrelated, and involves association of the enzyme from the bulk solution back to the DNA. The steps subsequent to the enzyme binding to the DNA limit the kinetics of binding to the specific site, and there are twice as many available sites for single cutting events as double cutting events. Therefore, $k_2 = k_1/2$ for completely unprocessive catalysis, resulting in $F_p = 1/3$ as the lower limit of processivity. For processive catalysis, $k_2 > k_1/2$ as the enzyme remains associated with the DNA following the single cut and preceding the second cut. Highly processive catalysis is associated with a relatively small accumulation of singly cut intermediates. For $k_2 >> k_1$, $F_p = 1$. However, endonucleases indiscriminately associate with either fragment generated after a DNA cleavage (14,27). Therefore only half of the cleavage events will result in the enzyme associating with the fragment with the second site. Given the lower limit of processivity, this means maximal processivity for endonucleases is $F_p = 0.66$ not 1.

Figure 3.2. ENase processivity assay. A) Non-denaturing PAGE gel showing the separation of products of a flourescien labeled two-site symmetric substrate with specific sites for either EcoRI or EcoRV. B) Uc, Sc', and Dc* are further processed (text) revealing uncut (Uc), singly cut (Sc), and doubly cut (Dc) DNA, which is the basis for the k_1, k_2, and ultimately processivity (F_p). C) Plot of Sc and Dc from gel in A.



3.3.2 EcoRI and EcoRV display different intersite processivity dependencies

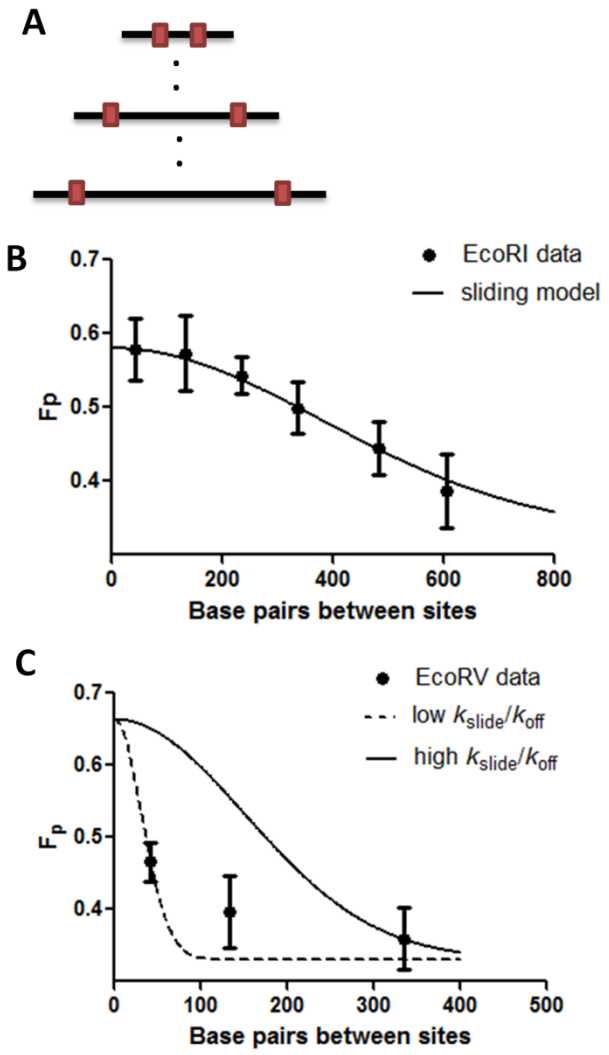
How processivity changes with respect to changes in intersite distance is the basis for elucidating translocation mechanisms (Equations 1,2). Here we use a series of two site substrates with flanking sequences held constant at 115 and 119 bp and the spacing between sites ranging from 42 to 605 bp for EcoRI (Figure 3.3A). We observe a processivity value of F_p = 0.58 +/- 0.03 with 42 bp between the two sites, meaning that EcoRI is largely processive with this substrate as processivity is maximally F_p = 0.66 (14,27). Importantly,

the inclusion of 65 mM NaCl is responsible for the slight lowering of processivity from the theoretical maximum, consistent with prior work (27). Furthermore, processivity remains largely unchanged with intersite distances up to 134 bp, and then only gradually decreases with further increases in intersite distances. EcoRI is processive up to ~600 bp.

Based on our results from Figure 3.3B, we extrapolate that EcoRI has a theoretical processivity value of ~0.58 when the two sites are directly next to each other (separated by 0 bp) under these conditions, and Equation 1 is normalized to reflect this, resulting in Equation 9. Under lower salt levels, processivity is slightly higher, approaching the theoretical maximum (below). Equation 9 also accounts for the fact that the maximum level of processivity for an endonuclease is lowered by the random partitioning between the products of the first cleavage event (see above), overall defining the range of our processivity assay under these experimental conditions at F_p = 0.33 - 0.58.

$$F_p = [(k_{slide} / (k_{slide} + k_{off}))^{n\wedge 2}]/4 + 1/3 \quad (9)$$

Figure 3.3. Processivity data for EcoRI ENase and EcoRV ENase shows different intersite spacing dependencies. A) Schematic of substrates with variation of intersite distances. B) Processivity data for EcoRI includes 1nM enzyme, 400nM DNA, 5 mM $MgCl_2$, 65mM NaCl, and buffer at 37 degrees. Processivity trend is fit to the sliding model (equation 1), where each sliding step is orders of magnitude more probable than a dissociation step (k_{slide}/k_{off} of 290,000:1). C) Processivity data for EcoRV includes 1.6 nM enzyme, 400nM DNA, 10 mM $MgCl_2$, 28mM NaCl, and buffer at 37 degrees. Processivity decreases more sharply as a function of the spacing between the two sites than in A. Fit to a higher k_{slide}/k_{off} of 45,000:1 (solid line) and a lower k_{slide}/k_{off} of 2,100:1 (dashed line). The processivity trend cannot fit to the sliding model, and the data is therefore consistent with a combination of sliding and hopping mechanisms.



The processivity data set fits the sliding model remarkably well (Figure 3.3B) with extensive sliding vs dissociation kinetics, requiring an extremely high k_{slide}/k_{off} ratio of

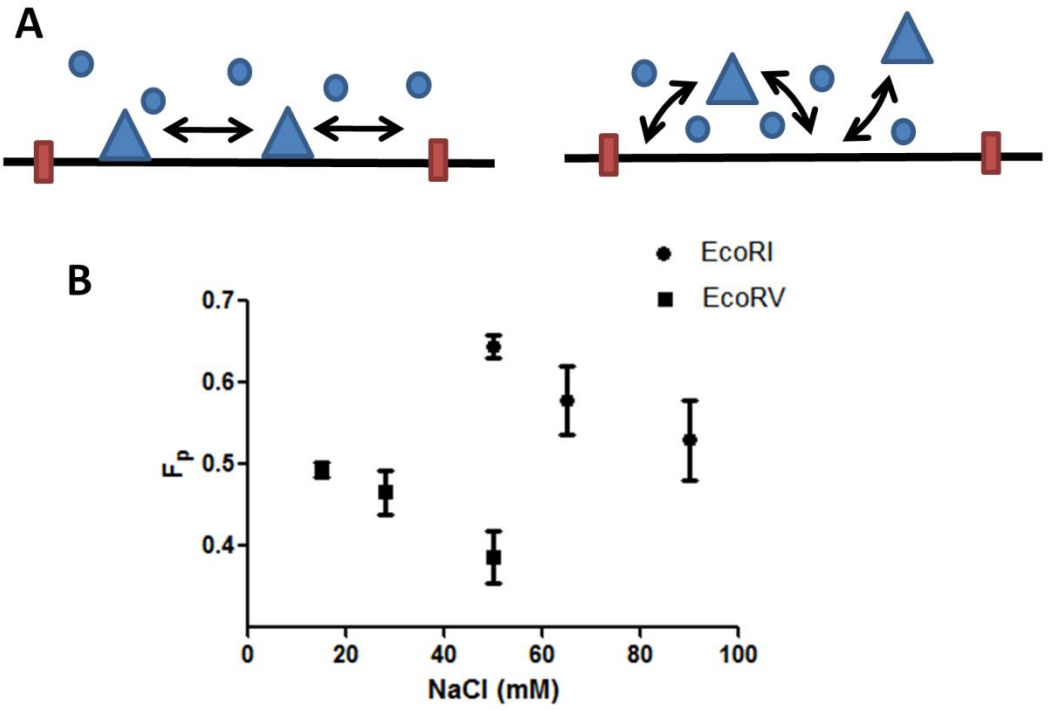
~290,000:1. Others have proposed that such a high ratio is unlikely and therefore invoke that hopping makes a significant contribution to the translocation mechanism (8,14). However, the excellent fit to the sliding model suggests this is not the case. We next applied the same measure of processivity to explore the translocation mechanism used by EcoRV, which has been extensively studied and our intention here was to essentially replicate the results of Stanford *et al.* with our assay (Figure 3.3C) (14). Importantly, many prior studies of EcoRI and EcoRV and other processive nucleic acid modifying enzymes have relied on different assays, making direct comparisons between related enzymes difficult. Here the series of intersite distances span from 42 to 335 bp. The salt concentration was lowered to 28mM to maximize the processivity values (below). The most obvious distinction between the two enzymes is that processivity decreases when increasing the site spacing from 42 bp to 134 bp for EcoRV, while the processivity remains essentially unchanged for EcoRI in this intersite distance range (Figure 3.3B). Importantly, the processivity trend observed for EcoRV cannot be reconciled solely with the sliding model, either using high or low k_{slide}/k_{off} ratios (Figure 3.3C). This suggests that EcoRV uses a combination of sliding and hopping, consistent with prior results (14). Also, a significant reliance on an intersegmental hopping mechanism would result in processivity increases starting at ~300 bp separation (17), which is not the case here or in prior work using longer intersite spacings (14).

3.3.3 The greater salt dependence by EcoRV than EcoRI suggests distinctive translocation mechanisms

The influence of salt concentration on processivity reveals important details concerning translocation mechanisms (15). Salt screens the interaction between the negatively charged DNA polymer and the positively charged DNA binding cleft on the protein. Importantly,

ENase processivity measurements determine how likely a second site is cut following an initial cutting event, limiting the read out of salt's impact only to the protein moving between the two sites. Decreases in processivity associated with increases in salt concentration more strongly implicate a hopping mechanism, as salt disrupts the ability of the protein to reassociate with the DNA as it hops (Figure 3.4A). During sliding, in contrast, the protein remains associated with the DNA throughout the translocation, resulting in an attenuated response to salt than in hopping. We challenged both EcoRI and EcoRV with increasing salt (NaCl) using a substrate with 42 bp between the sites.

Figure 3.4. Differential salt dependencies for EcoRI and EcoRV. A) Salt (circles) effects the processivity of a hopping enzyme (right) more dramatically than a sliding one (left). B) Processivity assays with different amounts of salt for a substrates with 42 bp between the sites. EcoRV: 15, 28, 50mM NaCl; EcoRI: 50, 65, 90mM NaCl. EcoRV's processivity is more sensitive to salt, consistent with a hopping mechanism. EcoRI's processivity is less sensitive to salt, which is consistent with a sliding mechanism. EcoRI is more likely to closely interrogate short distances along DNA under conditions of high salt (*in vivo*) than EcoRV.



EcoRI's processivity under low salt concentrations (F_p= 0.64 ± 0.01) (Figure 3.4B) is near the maximum level (F_p = 0.66), consistent with the defined limitations of our assay (see above). (This further corroborates the legitimacy of Eq. 9.) EcoRV's processivity does not extrapolate to the maximum level at low salt, suggesting that unlike EcoRI, intersite spacings shorter than those investigated here might reveal higher processivity values, further suggesting that EcoRV's sliding is less extensive than EcoRI's. The greater sensitivity to salt concentration by EcoRV is consistent with EcoRV being dissociated from the DNA for longer periods of time, presumably due to a greater reliance on a hopping mechanism. This is further evidence that the two ENases use distinct facilitated diffusion mechanisms. Our results suggest that EcoRV will show poor processivity and is therefore unlikely to closely interrogate short segments of DNA under physiological (100mM NaCl) conditions, as has been shown in prior work (14). Interestingly,, Lomhalt *et al.* showed that EcoRV undergoes intersegmental hopping by DNA looping at 100 mM NaCl (but not at 25 mM or 150 mM NaCl) (19). The Lomhalt *et al.* report contrasts with Gowers *et al.* (18), who initially showed that movement between plasmid concatamers by the same mechanism is only possible under low salt conditions, and does not occur when NaCl is ≥ 100mM. We sought to further probe the conditions and extent which EcoRV (and EcoRI) utilizes intersegmental hopping.

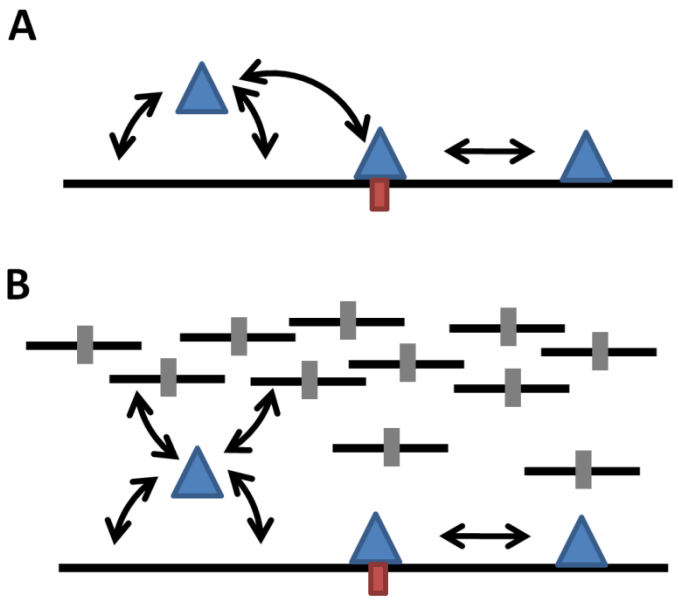
3.3.4 Chase Assay: EcoRV utilizes intersegmental hopping to a greater extent than EcoRI

Intersegmental hopping involves protein movement between two DNA segments that come into sufficient proximity (introduction). Importantly, we distinguish this from a similar process involving a protein or protein complex that simultaneously binds the two DNA

segments (intersegmental transfer).This simultaneous binding is not possible for the two ENases studied here (17). We use a chase assay to probe if EcoRV (and EcoRI) manifest intersegmental hopping under low (6 mM) and/or high (100 mM) NaCl conditions.

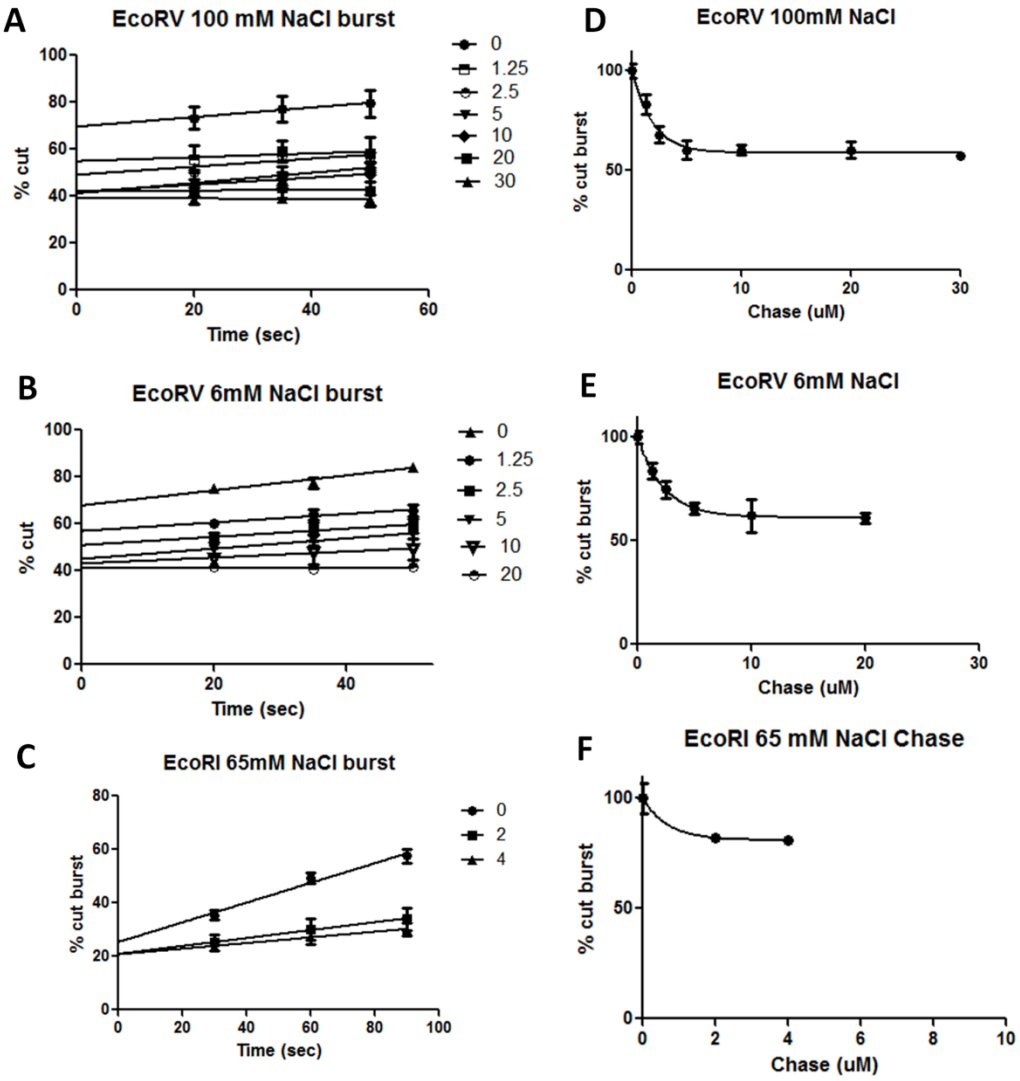
This assay relies on the fact that the enzyme incubated with the DNA generates cleaved product much faster than during subsequent steady-state catalytic cycles, where the enzyme releases product, returns to bulk solution, and cuts again (31). This generates a burst of product formation resulting from the use of relatively high enzyme concentrations combined with a ~1.5 fold excess of DNA (400 nM). EcoRV is first incubated on a single site substrate in the absence of metal, and cleavage is initiated by Mg^{2+} addition with various amounts of unlabeled competitor DNA chase. In the absence of divalent metal, EcoRV binds to specific DNA only ~5 fold tighter than non-specific DNA (32), and previous reports estimate even less discrimination (33). Therefore, prior to initiation by Mg^{2+}, EcoRV exists as three dynamic populations: bound to non-specific DNA surrounding the site, bound to the specific site, and protein which has dissociated from non-specific or specific DNA (Figure 3.5). Our detected burst level in the absence of chase indicates very little enzyme dissociates completely away from the DNA (Figure 3.6A, B). The population of enzyme that is not bound to the site but can return to it will do so by hopping or sliding (Figure 3.5).

Figure 3.5. Schematic of the chase assay. Enzyme (triangle) is in multiple dynamic populations bound to a single site (rectangle) substrate in the absence of metal or chase, poised for catalysis: 1) it remains bound to the site; 2) it can hop or slide to the site; 3) it leaves the DNA (not shown). A) The initiation of cleavage in the absence of chase allows the enzyme to cut, which includes hopping or sliding to the site. B) The initiation of cleavage with excess chase. Sufficient quantities of chase will capture the population of hopping enzymes by way intersegmental hopping.



The extent of inhibition by the chase is dependent on the fraction of the protein which hops during the initial phase of the reaction. The addition of the chase DNA should capture the fraction of enzyme which has separated from the DNA, likely during a hop, but in the absence of chase, returns and generates product. Thus, the extent to which the chase DNA decreases the burst magnitude is directly related to the fraction of protein which dissociates but normally reassociates (34). An enzyme that hops but cannot undergo intersegmental hopping will be unaffected by the chase.

Figure 3.6. EcoRV undergoes intersegmental hopping under low (6mM) and high (100mM) NaCl. A) 240nM EcoRV is incubated on 400nM DNA, with buffer at 4 degrees and 100mM NaCl. Initiation with Mg^{2+} and chase (0-30μM) causes a burst of product formation commensurate with the amount of enzyme, followed by slower steady state cleavage. Chase decreases the burst level. B) Burst assay with EcoRV and 6mM NaCl. Initiation is with Mg^{2+} and chase (0-20μM). C) Burst assay with 100nM EcoRI incubated on 400nM DNA, with buffer at 4 degrees and 65mM NaCl. D) Burst levels in A are depicted as a function of chase. This shows a clear plateau, signifying a portion of the enzyme population can be captured by the chase by undergoing intersegmental hopping. E) Burst levels in B are depicted as a function of chase, with a similar outcome as D. F) Bursts levels in C are depicted as a function of chase. Most of the enzyme is not captured by chase through intersegmental hopping.

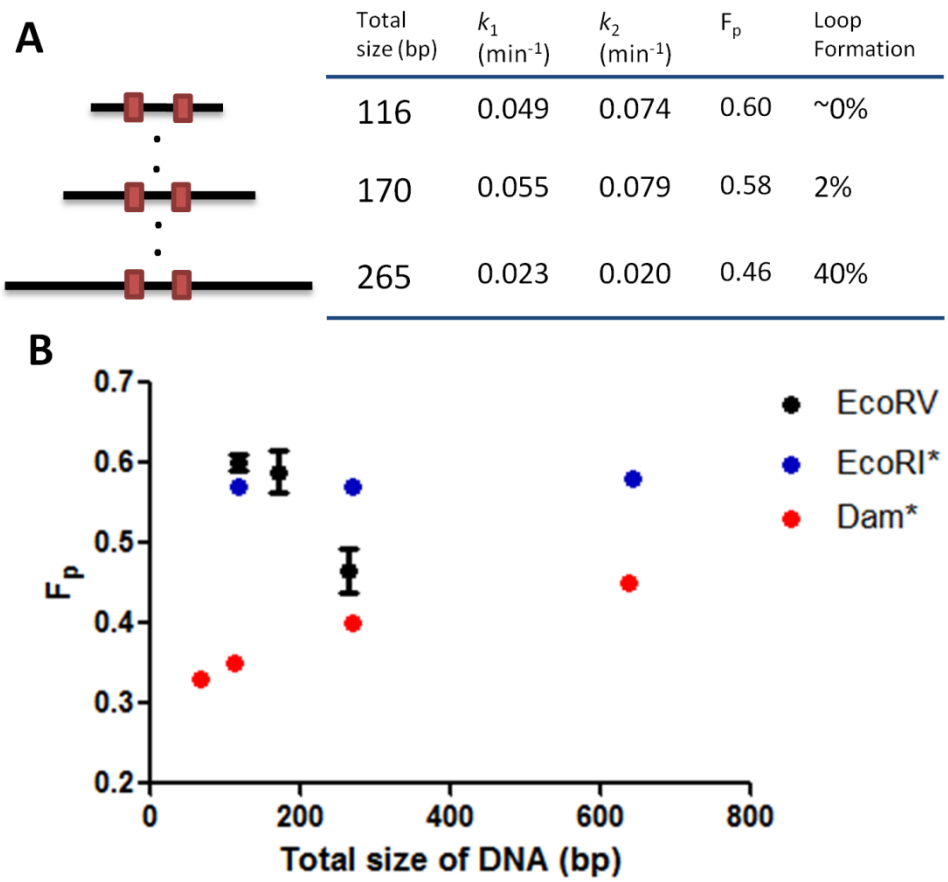


In comparison to the burst level with no chase, a five-fold excess of chase DNA causes a clear decrease in the burst magnitude, consistent with a portion of the enzyme being captured during intersegmental hopping events (Figure 3.6 D,E), either under low or high salt levels. Importantly, the effect of the chase on the burst plateaus, meaning the chase can only capture roughly 40% of the enzyme population, suggesting that the remaining enzyme population is poised on the specific site or able to move undisturbed (likely by sliding) from the non-specific sites to the specific site during this time period. Our interpretation of the partitioning of mechanisms as revealed by saturation levels of chase is similar to that of Stivers and co-workers (35).

Under similar conditions, EcoRI's burst is less affected by chase than EcoRV's (Figure 3.6F), consistent with a smaller fraction of the enzyme (~20%) being captured by intersegmental hopping, and prior similar EcoRI chase binding assays (15,36). This may be due in part to EcoRI's ~1,000 fold discrimination for specific over non-specific DNA (37), or its heavy reliance on sliding. A similar yet non-catalytic assay used by Siderova *et al.* (34) also detects intersegmental hopping for EcoRI. Importantly, this chase assay is useful in determining whether an enzyme can utilize intersegmental hopping (similarly to others using single site substrates) (18,19,34), but is uninformative to the relative extent other possible translocation mechanism are employed. EcoRV is capable of intersegmental hopping at salt concentrations where it is unlikely to be processive, suggesting this mechanism can contribute to site finding *in vivo*.

3.3. 5 Evidence for EcoRV reliance on intersegmental hopping: systematic variation in flanking DNA length

Figure 3.7. Processivity trends with changes to flanking DNA amounts. A) Processivity data for EcoRV, black dots. Loop formation probability (%) is based on the looping equation (24), where DNA under 200 bp is unlikely to undergo looping. Here, DNA >500bp have a 100% loop formation probability. Loss of processivity is consistent with intersegmental hopping "distracting" from sliding or hopping. B) Processivity data plot. EcoRI does not utilize looping and its processivity trend is not altered* within this set of substrates. Dam can only utilize intersegmental hopping and its processivity slightly increases*. *replot from Pollak *et al*. (Chapter 2).



The processivity trend from a series of substrates where the intersite spacing remains the same (42 bp), but the flanking DNA length changes, can be informative regarding translocation mechanism (Figure 3.7A) (17). For example, we recently showed that EcoRI's processivity is insensitive to increases in the amounts of flanking DNA (Figure 3.7B, chapter 2) (17). This lack of responsiveness is consistent with Eq. 1 (which only depends on

the amount of DNA between the sites) and therefore is further suggestive of a translocation mechanism that relies significantly on sliding (17). Importantly the size regime of the total length of the substrates coincides with a dynamic range of DNA loop formation probability, where DNA under ~200 bp won't loop but is increasingly likely to do so between 200-600 bp.

Unlike EcoRI, EcoRV's processivity is modulated with changes in flanking DNA lengths. The two shorter substrates (116 and 170 total bp size, Figure 3.7) are significantly more processive than the longer substrate (265 total bp size, Figure 3.7, replot from Figure 3.3). A compelling explanation for this trend involves the fact that intersegmental hopping proceeds though DNA loops on DNAs >200 bp (17). The smaller two substrates are therefore highly unlikely to utilize intersegmental hopping due to their size, and therefore the enzyme can only slide or hop, which it appears to do quite efficiently given the high processivity values (Figure 3.7A). The longer substrate is interestingly less processive. Since the sites are only 42 bp apart, intersegmental hopping may represent a distraction, guiding the enzyme away from sliding or hopping to the adjacent second site and instead transferring it to a distal DNA segment. This notion is corroborated considering the individual rate constants of the data set, where the longer substrate is slower than the two smaller ones (Figure 3.7A). As the search for the sites following the association between the enzyme and the DNA is rate limiting, movement between loops slows the reaction down. Importantly, the 265 bp substrate follows the expected trends from Figures 3.3 and 3.4, and therefore is unlikely to be an outlier.

3.4 Discussion

Facilitated diffusion on DNA is a ubiquitous cellular process employed by diverse classes of proteins, including transcription factors, DNA methyltransferases, restriction endonucleases, and DNA repair enzymes. That these proteins passively locate their largely undisclosed sites of action which can differ from non-target sites by a few atoms under biological timescales is impressive, yet not fully understood. While sliding, hopping, and intersegmental hopping mechanisms have been identified and characterized, examples of how these mechanisms are coupled to biological processes are seldom considered. Here, we continue to explore if particular proteins utilize different combinations of these mechanisms to accommodate their cellular function(s). We previously showed that Dam primarily uses intersegmental hopping in comparison to sliding and hopping mechanisms, which supports the necessary large movements each enzyme must make along the genome. In contrast, repair enzymes, such as UNG, are effective at moving short distances along DNA using hopping and sliding, but are poorly processive when sites are separated by more than ~100 bp. Repair enzymes have high cellular copy numbers (~10^5) and excise damaged bases that are very rare (1 every 10^6 bp) (12,15) Therefore *in vivo* processive catalysis is unlikely, and a scenario where each enzyme is confined to shorter stretches of genomic DNA seems probable. Furthermore, Dam readily skips sites (methylation marks separated by 1kb are sufficient for mismatch repair while each site is typically separated by only ~250 bp), while the repair of each target by UNG is more critical, and rigorous and repetitive searching of local regions is necessary. Biological context is clearly connected to the combination of translocation mechanisms a particular protein will use, and our study here reveals this for two closely related enzymes.

EcoRI and EcoRV display distinct translocation properties, and we are interested in whether this helps understand the enormous biological investment in the diversity and co-existence of distinct Type II systems. We provide several lines of evidence that EcoRI employs an extreme sliding mechanism, where it can make ~290,000 single bp movements within a single binding event before dissociation (Figure 3.3B). The excellent fit of our processivity data to the theoretical sliding model is the initial basis for this conclusion. Importantly, our EcoRI processivity results are consistent with several prior reports. For example, Modrich and colleagues demonstrated that EcoRI processivity is only slightly less when two sites are 377 bp apart compared to a 51 bp separation (27). Extrapolations by Rau and coworkers suggest that EcoRI uses a sliding mechanism and is capable of sliding 400 bp on average prior to dissociation (28).

In contrast, the EcoRV processivity data does not fit the sliding model, consistent with results from Stanford *et al.*(14). This precludes us from estimating a "sliding distance" for EcoRV, but is consistent with a mechanism involving a blend of sliding and hopping (Figure 3.3C). The minor dependence on salt concentration shown by EcoRI in comparison to EcoRV is further consistent with the former being more reliant on sliding, as each movement does not involve microscopic dissociation-reassociation steps as is seen in hopping. EcoRV's lack of processivity with short site spacing under high salt concentrations (Figure 3.4) is suggestive that EcoRV is unlikely to continuously interrogate short segments of DNA *in vivo* (~100 mM NaCl) (14). As cellular ionic strength conditions fluctuate up to 100 fold to respond to environmental changes (38), enzymes that undergo the same task but respond differently to salt are likely to be of benefit to the cell. Our chase DNA experiments (Figure 3.6) show that EcoRV is more likely to fluidly move between distal DNA segments,

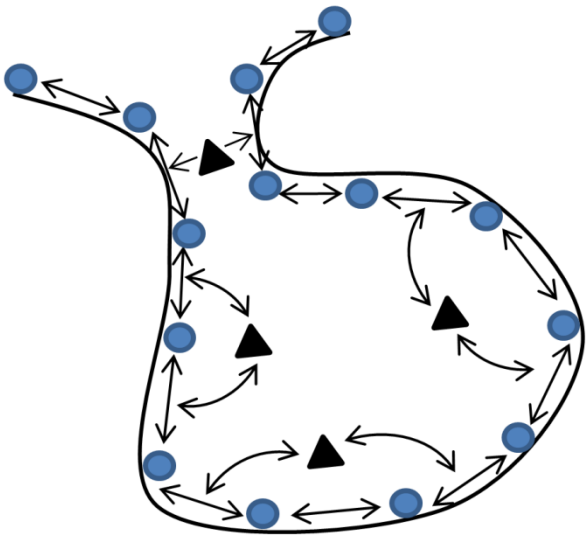
thereby expanding the breadth of DNA it can search. This is most dramatically highlighted considering Figure 3.7, where EcoRV is less effective in processively cleaving shortly spaced sites when the size of the DNA increases. Unlike Dam and EcoRI, which primarily rely on intersegmental hopping and sliding respectively, EcoRV doesn't appear to dominatly use either, and instead likely relies on a blend of mechanisms.

Our interpretation of the translocation mechanism used by EcoRV in Figure 3.7 is reconciled by comparisons with prior EcoRI and Dam data sets (17). If sliding dominates translocation then EcoRV's processivity will be insensitive to changes in flanks; therefore, sliding dominates for EcoRI but not for EcoRV. Importantly, our result is consistent with EcoRV utilizing intersegmental hopping to a greater extent than EcoRI. This variation (Figure 3.7) of the standard processivity assay (Figure 3.3) succeeds in partitioning intersegmental hopping from other potential mechanisms, and suggests that EcoRV uses intersegmental hopping in relative tandem with sliding and hopping mechanisms, without one being especially dominant. In contrast EcoRI dominantly uses sliding. Reconciliation with prior Dam data is more convoluted: since Dam is nearly incapable of sliding/hopping, the addition of flanking DNA to provide looping actually increases its processivity. However, we imagine the path resulting in processive catalysis is quite circuitous here. In contrast, EcoRV uses sliding/hopping to a greater extent than Dam. Therefore, intersegmental hopping by flanking nonspecific DNA confounds processivity for EcoRV for these substrates considering that traveling the short distance between sites is much more likely to be productive using sliding/hopping. Previous reports show that intersegmental hopping further assists sliding/hopping mechanisms in site locating for EcoRV; however,

these reports concern long (several kb) substrates, where intersegmental hopping is likely to be helpful (18,19). In contrast, Figure 3.7 concerns movement between only 42 bp.

Relative protein expression levels may help explain why EcoRI and EcoRV have different translocation properties but still can individually effectively cleave phage DNA. In *E. coli*, EcoRI has ~ 1000 dimer copies (39,40), while EcoRV has only 100 dimer copies (41). Due to its sliding mechanism, each EcoRI enzyme is confined to a smaller region of DNA; however, the abundance of protein copies extends its occupancy along DNA (Figure 3.8). While rarer, each EcoRV enzyme can extend its footprint to much more distant regions of DNA. These two distinct translocation mechanisms may function synergistically to ensure phage defense. Importantly, the drawback of one of these translocation mechanisms is mitigated by the other, creating a robust phage restriction strategy.

Figure 3.8. Relative protein levels explain translocation diversity. EcoRI's abundance of cellular protein levels in comparison to EcoRV helps explain why both can effectively cleave phage DNA. EcoRI (circles) redundantly slides over smaller regions, but large cellular copy numbers allow it to search more of the DNA. EcoRV's (triangles) ability to undergo hopping and intersegmental hopping causes site skipping but affords a broader search of the DNA.



There are other distinctions between these two enzymes that make their differing translocation mechanisms seem more reasonable. EcoRI does not bend the DNA at its site, while EcoRV does (29). As protein-mediated DNA bending can facilitate DNA looping (42), this may be related to enhancing the ability of EcoRV to carry out intersegmental hopping. EcoRV leaves "blunt" ends while those of EcoRI are "sticky," leaving 4 overhanging nucleotides. These overhanging bases may facilitate the enzyme to stay more closely positioned to its cutting site following catalysis as DNA binding proteins readily bind to sticky ends. In contrast EcoRV leaves its site more readily following cutting, allowing it to be intercepted by proximal looping DNA. The blunt and sticky end distinction between EcoRI and EcoRV is also strong evidence that these enzymes evolved in different organisms and therefore may have evolved separate translocation mechanisms (6). The extent of the distinction between these two enzymes may be surprising to those familiar with their similarities: their recognition sites are nearly identical, the proteins are similar in size; they use the same cofactor and maintain the same oligomeric state for catalysis (6). Structural classification suggests evolutionary divergence separates blunt end and sticky end generating ENAses (43,44). Interestingly, unlike their methyltransferase partners, there is very little homology or correlation between structure and function between ENases other than the conserved (D/E)XK catalytic motif (45). Unfortunately, a structural basis for predicting translocation mechanism does not exist for these "saddle proteins" (46), other than in examples of proteins with positively charged, disordered, DNA binding tails (22).

An appreciation of ENase translocation mechanism diversity may help to explain why 80% of bacteria have multiple R-M systems. Given the high fitness cost of maintaining R-M

systems (1), acquiring several may seem superfluous. However, certain strains of bacteria contain as many as 16 R-M systems (47,48). Phage anti-restriction mechanisms, other bacterial phage defense systems distinct from R-M systems (introduction), and the evolutionary advantage of acquiring foreign DNA, all highlight the intricacies and extent of the arms race between bacteria and invading foreign nucleic acids (1,5). Having several distinct R-M systems in a single cell is in part correlated with the propensity of R-M systems to act as "selfish elements" spreading through horizontal gene transfer, where their removal is avoided as it results in post-segregational killing (40). Another explanation of replicate R-M systems relies on their involvement in the evolution of new strains, where differing recognition specificities can segregate species into distinct variants.(1,47). Interestingly, bacteria lacking RecBCD, which effectively cleaves invading phage DNA, contain more R-M systems (47), suggesting multiple R-M systems can have an additive effect on phage defense. Overall the prevalence and functional consequences of the redundancy of type II R-M systems remains a mystery (47), which is also the case for the related type II toxin-antitoxin systems (49). We suggest that a plausible explanation for the existence of several restriction systems in a single organism is their mechanistic differences, and that translocation mechanisms may contribute to this. Instead of redundancy, distinct site finding strategies likely confer effective phage restriction and may contribute to the other functions of R-M systems.

3.5 Materials and methods

Processivity Assays: EcoRI was diluted into the following storage buffer: 10mM KPO_4 300mM NaCl,0.1mM EDTA, 200ug/mL BSA, 10mM DTT, 0.15% Triton X, and 50% Glycerol. Reactions were done in the EcoRI reaction buffer: (20 mM Tris pH 7.4, 5 mM

$MgCl_2$, 0.2 mM EDTA, 0.05 mg/ml BSA) with 400nM DNA, and a total of 65 mM NaCl at 37 degrees. Reactions were initiated by the addition of 1nM EcoRI and time points were quenched in 50mM EDTA.

EcoRV was diluted into the following storage buffer: 10mM Tris-HCl pH = 7.5, 50mM NaCl, 1mM DTT, 0.1mM EDTA, 200ug/mL BSA, and 50% Glycerol. Reactions were done in EcoRV reaction buffer: (50mM Tris-HCl, pH 7.5, 10mM $MgCl_2$, 0.25mM EDTA, 50ug/mL BSA) with 400nM DNA, and a total of 28 mM NaCl at 37 degrees. Reactions were initiated by the addition of 1.6 nM EcoRV and time points were quenched in 50mM EDTA.

Salt study: Reactions were carried out exactly as above with NaCl levels modulated. For EcoRI: 50, 65, and 90mM NaCl were used. For EcoRV; 15, 28, and 50mM NaCl were used.

Chase Assays: 240nM EcoRV was incubated on 400nM DNA, with buffer (above) at 4 degrees for 30 minutes at the designated salt concentrations. The reaction was initiated with a mixture of 10mM Mg^{2+} and chase (0-30μM) as indicated. Reactions were quenched in 50mM EDTA. 100nM EcoRI was incubated on 400nM DNA, with buffer (above) at 4 degrees for 30 minutes at 65mM NaCl. The reaction was initiated with a mixture of 10mM Mg^{2+} and chase (0-4μM) as indicated. Reactions were quenched in 50mM EDTA.

Enzymes: EcoRI was purchased from NEB. Purified EcoRV was kindly provided for by Dr. John Perona. The purification protocol can be seen in Hiller *et. al* (29).

Data Analysis: The reaction products were analyzed using PAGE (20%-5% (depending on substrate length) 29:1 acrylamide: bisacrylamide) at 300V for 1-3 hours. Gels were scanned on a Typhoon Phosphoimager (GE). Densitometry was done on the provided

software (GE). The processivity values were derived from a least squares fitting using Microsoft excel.

DNA substrates: Overview: All restriction endonucleases for cloning were purchased from NEB. All synthetic DNA substrates and primers were purchased from IDT and were re-suspended in TE buffer (10mM Tris, pH 7.5, 1mM EDTA). When necessary, DNA was annealed with its reverse complement in a 1:1 mixture for 5 minutes at 95 degrees, and allowed to slow cool to room temperature (~5 hours). Annealing was verified by PAGE. The production of several PCR amplicons used the following strategy: A synthetic oligonucleotide with two EcoRI sites (below), and two restriction sites between the EcoRI sites was cloned into plasmid pBR322 (NEB), making a new plasmid, pBR322GAATTC. The same strategy was used to generate the EcoRV substrates making another plasmid, pBR322GATATC. Several spacers (below) were generated by PCR (from a different region of the pBR322 plasmid) and were cloned into pBR322GAATTC and pBR322GATATC, generating several plasmids with different distances between the GAATC and GATATC sites. Amplicons were generated by PCR using a single set of primers for substrates used in Fig. 3.3, but different sets in Fig 3.7. Amplicons were purified using a PCR clean-up kit (Qiagen). Substrates used in specific figures are described below.

Figure 3.3B: The following substrate was annealed then cloned into the digested plasmid pBR322 at the EcoRI and HindIII sites, 5'-agcttggcattgaattcta*ccatgg*gcatactaccagct*cccggg*tcggctcgatcaccg-3' ; 5'-aattcggtgatcgagccga*cccggg*agctggtagtatgc*ccatgg*tagaattcaatgcca-3', making pBR322GAATTC. The cloned, synthetic insert had additional cloning sites within it: XmaI and NcoI (italicized). These sites were used to insert PCR purified spacers between the two

GAATTC site(s) (underlined). Upon PCR amplification (below), the spacers were digested with XmaI and NcoI to generate overhangs. The spacers were generated by PCR from the plasmid pBR322 with restriction sites using the same forward primer: 5'- att *cccggg* ggctaccctgtggaacacct - 3', with different reverse primers for each sized spacer, starting with substrate 3B-2:

Numbers refer to bp between sites, total bp size once the spacer is cloned into pBR322GAATTC

3B-1, 42, 270

(no spacer)

3B-2, 134, 364:

5'- taatccatggtactggaacgttgtgagggt-3'

3B-3, 234, 468:

5'- taatccatgggcctccgtgtaaggggatt-3'

3B-5, 482, 716:

5'- taatccatggcccggcatccgcttacagac-3'

3B-6, 605, 839:

5'- taatccatggcaatctgctctgatgccgca-3'

As a control, a spacer was constructed from a different region of the plasmid, using a different forward primer.

Forward: 5'-attcccgggatagttgcctgactccccgt-3'

3B-4, 335,569:

5'- taatccatggcgttgggaaccggagctgaa-3'

The amplicons with 115/119 base pairs flanking GAATTC sites were amplified from each plasmid using primers: (forward) 5'- gggttccgcgcacatttccc-3' and (reverse) 5'-(Fl) ccagggtgacggtgccgagg-3'.

Figure 3.3C: The same strategy and sets of primers to generate the spacers was used for the EcoRV substrates. The following substrate was annealed and cloned into digested pBR322 to generate pBR322GATATC 5'-aattcgatatcttcga*cccggg*agcagcatctcgtgagtagtgc*ccatgg*ttcgatatcca-3'; 5'-agcttggatatcgaa*ccatggg*cactactcacgagatgctgct*cccggg*tcgaagatatcg-3'. The same (select) set of spacers from above was used to generate the substrates for Figure 3.3C: 3B-2, 134 bp between sites, 364 bp total size; and 3B-4, 335 bp between sites, 569 bp total size.

The amplicons with 115/119 base pairs flanking GAATTC sites were amplified from each plasmid using primers: (forward) 5'- gggttccgcgcacatttccc-3' and (reverse) 5'-(Fl) ccagggtgacggtgccgagg-3'.

Figure 3.6: The following substrate was annealed and cloned into the digested plasmid pBR322GAATTC (for convenience) at the XmaI and NcoI sites: 5'-Ccgggtacaacttggatatcggttacacgcc 3'; 5'catgggcgtgtaaccgatatccaagttgtac-3' making a plasmid pBR322GAATTCrvchase. The same primers from above (the 115/119 set) were used to generate the amplicon for the EcoRV chase experiment.

The following substrate was annealed and used as the chase DNA for EcoRV reactions: 5'- Ccgggtacaacttggatatcggttacacgcccatg-3',5'-catgggcgtgtaaccgatatccaagttgtacccgg-3'

The same primers from above (the 115/119 set) were used to generate the amplicon for the EcoRI chase experiment using pBR322GATATC (for convenience).

The following substrate was annealed and used as the chase DNA for the EcoRI reactions: 5'- Ccgggtacaacttggaattcggttacacgcccatg-3',5'-catgggcgtgtaaccgaattccaagttgtacccgg-3'

Figure 3.7: pBR322GATATC was PCR amplified with a different sets of primers to adjust the spacings from GATATC to the ends of the DNA. The amplicon with 30 base pairs flanking the GATATC sites was amplified from the plasmid using primers: (forward) 5'-ggcgtatcacgaggcccttttcg -3', and (reverse) 5'-(Fl) cgttagcaatttaactgtgataaactaccgc -3'. The Amplicon with 60 base pairs flanking GATATC was amplified from the plasmid using primers: (forward) 5'- acattaacctataaaaataggcgtatcacg -3', and (reverse) 5'-(Fl) cggtgcctgactgcgttagcaa -3'.

3.6 References

1. Murk, I. & Kobayashi, I. (2014) To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res.* **42**, 70-86

2. Sorek, R., Kunin, V. & Hugenholtz, P. (2008) CRISPR – A widespread system that provides acquired resistence against phages in bacteria and archea. *Nat. Rev. Microbiol.* **6**, 181-186

3. Dillingham, M. S. & Kowalczykowski, S. C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **4**, 642- 671

4. Labrie, J. S., Samson, J. E. & Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317-327

5. Tock, M. R., & Dryden, D. T. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472

6. Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* **62**, 685–707

7. Berg, O. G., Winter, R. B. & von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929-6948

8. Halford, S. E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans*. **37**, 343-348

9. Halford, S. E. & Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res*. **32**, 3040-3052

10. Ghaemmaghami, S. *et al.* (2003) Global analysis of protein expression in yeast. *Nature* **425**, 737-741

11. Elf, J., Li, G. W. & Xie, X. S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194

12. Friedman, J. I. & Stivers, J. T. (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957-4967

13. Cai, L., Friedman, N. & Xie, S. X. (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* **440,** 358-362

14. Stanford, N. P., Szczelkun, M. D., Marko, J. F. & Halford,S.E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* **19**, 6546-6557

15. Hedglin, M. & O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. *ACS Chem. Biol.* **5**, 427-436

16. Porecha, R. H. & Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10791-10796

17. Pollak, A. J, Chin, A. T. Brown, F. L. H. & Reich, N.O. (2014) DNA Looping Provides for "Intersegmental Hopping" by Proteins: a Mechanism for Long-Range Site Localization. *J. Mol. Biol.* DOI: 10.1016/j.jmb.2014.08.002

18. Gowers, D. M. & Halford, S. E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO. J.* **22**, 1410-1418

19. van den Broek, B., Lomholt, M. A., Kalisch, S. M. J., Metzler, R. & Wuite, G. J. L. (2008) How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 1578-15742

20. Hedglin, M., Zhang, Y. & O'Brien, P. J. (2013) Isololating contributions from intersegmental transfer to DNA searching by Alkyladenine DNA Glycosylase. *J. Biol. Chem.* **288**, 24550-24559

21. Iwahara, J., Zweckstetter, M. & Clore, G. M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15062-15067

22. Vuzman, D., Azia, A. & Levy, Y. (2010) Searching DNA via a "monkey bar" mechanism: The significance of disordered tails. *J. Mol. Biol.* **396,** 674-684

23. Shore, D., Langowski, J. & Baldwin, R. L. (1981) DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4833-4837

24. Ringrose, L. & Chabanis, S. (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO J.* **18**, 6630-6641

25. Wright, D. J., Jack, W. E. & Modrich, P. (1999) The kinetic mechanism of the EcoRI endonuclease. *J. Biol. Chem.* **274**, 31896-31902

26. Jeltsch, A., Alves, J., Wolfes, H., Maass, G. & Pingoud, A. (1994) Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA. *Biochemistry* **33**, 10215–10219

27. Terry, B. J., Jack, W. E. & Modrich,P. (1985) Facilitated diffusion during catalysis by EcoRI endonucleases. Non-specific interactions in EcoRI catalysis. *J. Biol. Chem.* **260**, 13130-13137

28. Rau, D. C. & Sidorova, N. Y. (2010) Diffusion of the restriction nuclease EcoRI along DNA. *J. Mol. Biol.* **395**, 408-416

29. Hiller, D. A., Rodriguez, A. M. & Perona, J. J. (2005) Non cognate enzyme-DNA complex: structural and kinetic analysis of EcoRV endonucleases bound to the EcoRI recognition site GAATTC. *J. Mol. Biol.* **354**, 121-136

30. Fersht, A. (1998) *Structure and Mechanism in Protein Science*, W. H. Freeman, New York.

31. Baldwin, G. S., Sessions, R. B., Erskine, S. G. & Halford, S. E. (1999) DNA cleavage by the EcoRV restriction endonucleases: roles of divalent metal ion in specificity and catalysis. *J. Mol. Biol.* **288**, 87-103

32. Sidorova, N. Y., Muradymov, S. & Rau, D. C. (2011). Solution parameters modulating DNA binding specificity of the restriction endonuclease EcoRV. *FEBS J*. 15, 2713-2727

33. Taylor, J. D. & Halford, S. E. (1989) Discrimination between DNA sequences by the EcoRV restriction endonuclease. *Biochemistry* **28**, 6198–6207

34. Sidorova, N. Y., Scott, T. & Rau, D. C. (2013) DNA concentration dependent dissociation of EcoRI: direct transfer or reaction during hopping. *Biophys. J.* **104**, 1296-13039

35. Schonhoft, J. D. & Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA. *Nat. Chem. Biol.* **8**, 205-210

36. Pluciennik, A. & Modrich, P. (2007) Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12709-12713.

37. Sidorova, N. Y. & Rau, D. C. (1996) Differences in the water release for the binding of EcoRI to specific and nonspecific DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12272-12277.

38. Record, M. T., Courtney, S.E., Cray, S.D. & Guttman, H. J. (1998) Responses of E. coli to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends Biochem. Sci.* **23**, 143-148.

39. Modrich, P. & Zabel, D. (1976) EcoRI endonuclease. Physical and catalytic properties of the homogeneous enzyme. *J. Biol. Chem.* **251**, 5866–5874.

40. Ichige, A. & Kobayashi, I. (2005) Stability of EcoRI restriction-modification enzymes in vivo differentiates the EcoRI restriction-modification system from other postsegregational cell killing systems. *J. Bacteriol*. **187**, 6612-6621.

41. Bougueleret, L., Techini, M. L., Botterman, J. & Zabeau, M. (1985) Overproduction of the EcoRV endonuclease and methylase. *Nucleic Acids Res*. **13**, 3823−2839.

42. Hodges-Garcia, Y., Hagerman, P. J. & Pettijohn, D. E. (1989) DNA ring closure mediated by protein HU. *J. Biol. Chem.* **264**, 14621-14623.

43. Bujnicki, J. M. (2000) Phylogeny of the restriction endonuclease-like superfamily inferred from comparison of protein structures. *J. Mol. Evol.* **50**, 39–44.

44. Niv, M. Y., Ripoll, D. R., Vila, J. A., Liwo, A., Vanamee, E. S., Aggarwal, A. K., Weinstein, H. & Scheraga H.A. (2007)Topology of Type II REases revisited; structural classes and the common conserved core. *Nucleic Acids Res*. **35**, 2227-2237.

45. Kosinski, J., Feder, M. & Bujnicki, J. M. (2005) The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinforma* **6**,172.

46. Breyer, W. A. & Matthews, B. W., (2001) A structural basis for processivity. *Protein Sci.* **10**, 1699-1711.

47. Vasu, K. & Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* **77**, 53-72.

48. Stein, D. C., Gunn, J. S., Radlinska, M. & Piekarowicz, A. (1995) Restriction and modification systems of *Neisseria gonorrhoeae*. *Gene* **157**, 19–22.

49. Melderen, L. V. (2010) Toxin-antitoxin systems: why so many, what for? *Curr. Opin. Microbiol.* **13**, 781-785.

IV. The role of preexisting protein-DNA "roadblock" complexes in the search mechanisms by sequence-specific DNA binding proteins

4.1 Abstract

The genomes of all cells are intimately associated with proteins, which are important for compaction, scaffolding, and gene regulation. Here we show that pre-existing protein-DNA complexes (roadblocks) diminish and—interestingly—enhance the ability of particular sequence-specific proteins to move along DNA to locate their binding sites. We challenge the bacterial DNA adenine methyltransferase (Dam, recognizes 5'-GATC-3') with an Lrp (leucine-responsive regulatory protein)-DNA complex. The histone-like Lrp binds DNA non-specifically as an octamer. We observe an Lrp-dependent *increase* in processive (multiple modifications) methylation by Dam. These results can be explained by our prior demonstration that Dam moves over large regions (>300 bp) within a single DNA molecule using an "intersegmental hopping" mechanism. This mechanism involves the protein hopping between looped DNA segments. Lrp enhances Dam processivity through Lrp's ability to loop DNA, and therefore facilitate intersegmental hopping. Importantly this only occurs when the Dam sites are separated (by >134bp) such that they can be looped around Lrp. Intersegmental hopping may well be a general mechanism for proteins that navigate long distances along compacted DNA. Unlike Dam, EcoRI ENase (recognizes 5'-GAATTC-3') relies extensively on a sliding mechanism, and as expected, Lrp causes significant decreases in processivity. Furthermore, specific conformations of two tightly bound EcoRV ENase roadblocks also result in increases in Dam processivity. Overall, our results disfavor a reliance on the heavily touted sliding and hopping mechanisms for Dam, and instead

support intersegmental hopping. Our systematic use of protein roadblocks to investigate site location mechanisms provides a powerful strategy to differentiate between related mechanisms.

4.2 Introduction

The search for specific sites by DNA binding/modifying proteins is a ubiquitous biological phenomenon (1,2). Examples of biologically and bio-medically relevant proteins that move by diffusion (without ATP) along DNA to locate their inconspicuous sites of action include DNA methyltransferases (3,4), transcription factors (5,6), restriction endonucleases (7,8,9), and DNA repair enzymes (10,11). In general, nonspecific DNA overwhelmingly outnumbers specific sites, and this can be as little as ~100 fold and up to more than a million fold (10,12,13). Various mechanisms contribute to this facilitated diffusion process, where movement along non-specific DNA by proteins is critical for the labyrinthine site locating task, as originally reported in the 1970's (14) .

Interestingly, how proteins move along DNA given the litany of other DNA binding/bending proteins that crowd and manipulate DNA remains relatively experimentally unexplored. For example, *E. coli* leucine-responsive regulatory protein (Lrp) is one of over a dozen bacterial nucleoid-associated proteins (NAPs) whose overwhelming presence coats ~30% of the bacterial genome (15,16,17), and therefore certainly impacts facilitated diffusion. High cellular copy numbers and high affinities for non-specific DNA (18) allow NAPs to simultaneously control DNA structure and gene regulatory processes; Lrp, for example, controls ~10% of *E. coli* gene expression (17). The coverage of genomic DNA in eukaryotes (by histones) is much more extensive (17). One might intuit -- and it has been suggested (19,20) -- that these DNA-structure-manipulating-proteins will confound the site

finding process; however, we interestingly demonstrate here that this is not necessarily the case, as has been invoked by others (21-24).

In vitro processive catalysis studies are commonly used to deduce the details concerning translocation mechanisms, where the ability to carry out two modifications within a single binding event between the enzyme and the DNA is quantified (3,9-11). How processivity changes with changes in the spacings between the sites on the DNA is critical for identifying and discriminating between specific facilitated diffusion mechanisms (2). Sliding is one mechanism, where the protein makes a series of single base pair (bp) shifts to travel along the DNA, maintaining close contact with the DNA (9,10). Another mechanism, hopping, involves a series of dissociation-reassociation steps, allowing the protein to move larger distances (hops) during each step (9,10). For sliding, the drop in processivity corresponding with increases in the spacings between the two sites is sharper than for hopping (9,10). Both sliding and hopping rely on protein movement along the trajectory of the helix (Figure 4.1A), and are touted as the main mechanisms responsible for facilitated diffusion (1,11,22).

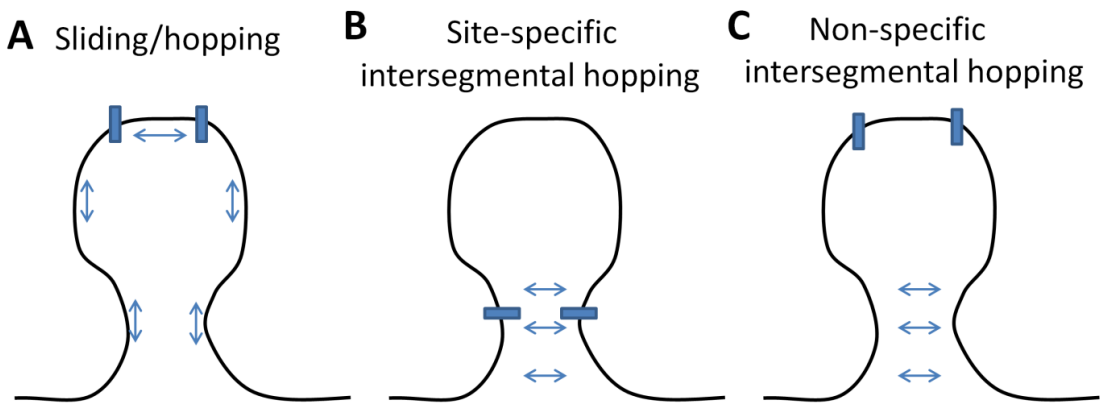
Several emerging lines of evidence suggest that translocation mechanisms other than sliding and hopping exist, involving large, "3-D" movements. For example, the ability of DNA to form loops, allowing distal (>300 bp) DNA segments from the same molecule to become proximal, provides proteins an alternative avenue for translocation and results in efficient long-range movements (Figure 4.1B,C) (3). The looping of DNA is dependent on its persistence length, where DNA segments less than 150 bp are rigid rods (25), and segments 200-600 bp are increasingly able to form loops (26,27). The protein moves between these looped DNA segments by what is likely a hop. Recent evidence for this mechanism has been reported for EcoRV ENase (EcoRV) (28,29), EcoRI ENase (EcoRI)

(30), alkyl adenine glycosylase (AAG) (31), and DNA adenine methyltransferase (Dam) (3). We refer to this mechanism as intersegmental hopping (3). Intersegmental hopping enhances processive catalysis for Dam when the two sites are far apart enough to be looped into proximity, called "site-specific intersegmental hopping," which optimally occurs with a ~500 bp separation between sites (Figure 4.1B). Intersegmental hopping also enhances processivity when the transition is between segments of nonspecific looping DNA (3,29); this "nonspecific intersegmental hopping" causes increases in processivity for Dam even when the two sites are close (Figure 4.1C), involving what is probably quite a circuitous path. Increases in processivity when the distances between or surrounding sites are increased (due to intersegmental hopping) to the best of our knowledge have only been demonstrated for Dam (3), but may well be a common strategy since many proteins are required to scan much larger regions of DNA than is typically studied *in vitro.* Intersegmental hopping is distinct from the well-studied intersegmental transfer mechanism (5,6,32), which necessitates the formation of a ternary complex with either a monomeric protein with two DNA binding sites, or a protein complex.

E. coli Dam is an orphan methyltransferase and uses AdoMet to methylate the N^6 position of adenine in 5'-GATC-3' sites (33). We recently demonstrated that Dam relies heavily on intersegmental hopping during processive catalysis, while not primarily using the more typically observed sliding or hopping mechanisms (3). However, its biological roles may help explain its translocation properties. Its participation in mismatch repair, involving the methylation of ~20,000 GATC, sites proceeds effectively with only ~100 Dam molecules per cell (13,34). Dam's inability to processively methylate closely spaced sites (3) is likely inconsequential, as not all GATC sites are likely required to be methylated for

mismatch repair to be effective (35,36), and skipping sites by traveling through DNA loops is ideal to efficiently cover the breath of the bacterial genome. Furthermore, Dam's function in epigenetic gene regulation involves only ~50 GATC sites that are clustered in certain regulatory regions (37) and are demonstrated to undergo non-processive catalysis. DNA binding/modifying proteins will utilize different combinations of the available translocation mechanisms, likely based on their biological context (3).

Figure 4.1. Facilitated Diffusion mechanisms. Arrows depict trajectory of protein movements along DNA. **A**) Well-studied sliding and hopping mechanisms involve movement along the trajectory of the DNA helix. **B**) Site specific intersegmental hopping takes advantage of when DNA loops involve sites that can be looped proximal in space. **C**) Nonspecific intersegmental hopping involves the movement between looped segments that lack sites. For Dam, this can cause increases in processivity even when sites are closely spaced. However, this mechanism is likely helpful for proteins that need to search for sites that can be quite rare.



Previous studies have utilized a "roadblock" approach to probe translocation mechanisms, where modulations of a protein's ability to travel between two sites and undergo processive catalysis is monitored with respect to the presence (and absence) of a pre-incubated protein roadblock (10,36,38). For example, O'Brien and colleagues concluded alkyl adenine glycosylase (AAG) utilizes hopping as a catalytically inactive EcoRI

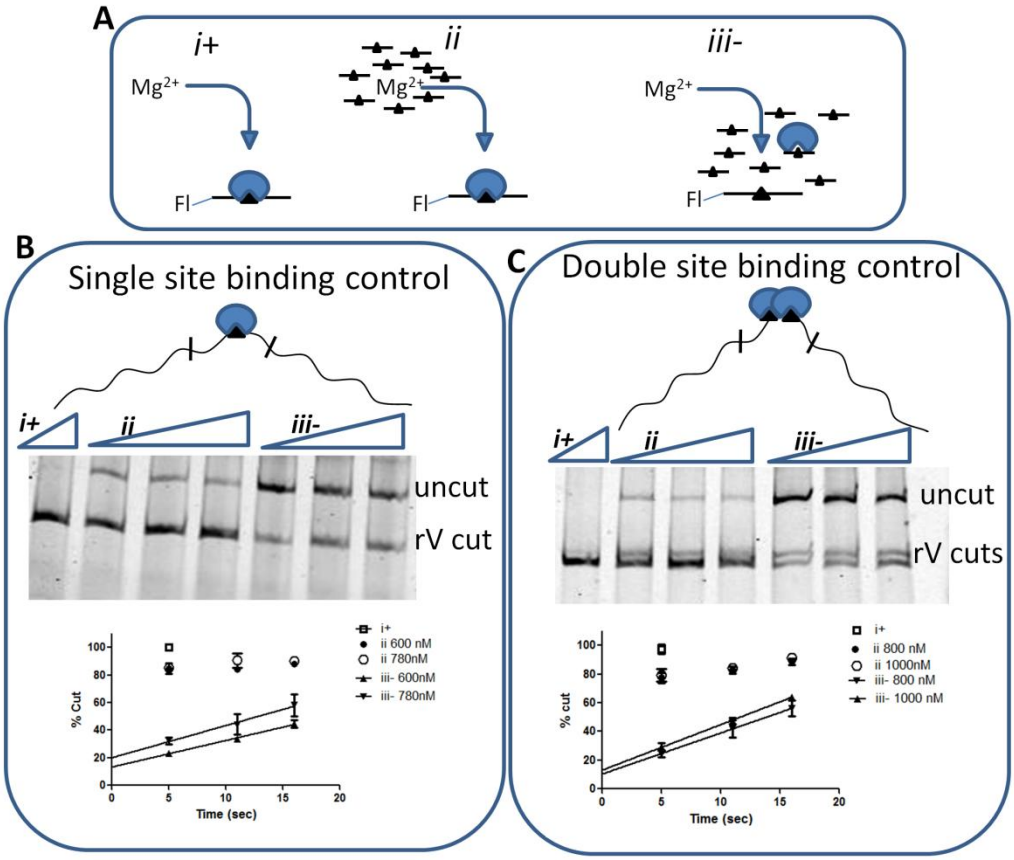
roadblock only partially disturbs the processive catalysis by AAG (10). Here we use an ensemble of highly stable complexes as well as transient complexes with two proteins (Dam and EcoRI) suggested to have very distinct translocation mechanisms (3). Furthermore, we employ a histone-like Lrp roadblock in the interest of probing a more biologically relevant situation for studies of facilitated diffusion. Our systematic approach provides a powerful way to distinguish between translocation mechanisms, and provides further evidence in support of the relevance of the intersegmental hopping mechanism.

4.3 Results

4.3.1 Demonstration of site-specific EcoRV occupancy as a roadblock

We employ EcoRV as a roadblock. EcoRV is a type II restriction endonuclease that recognizes and cuts both strands of 5'-GATATC-3' as a dimer upon addition of the Mg^{2+} cofactor (39,40). EcoRV binds its sites several orders of magnitude tighter than non-specific DNA in the presence of Ca^{2+} (39,41), which forms a stable non-catalytic complex with a very slow off rate (41), allowing it to maintain its position throughout the time course of our reaction. We are also interested in EcoRV's ability to bend DNA ~50° (39), as this expands the scope of possible DNA configurations to explore with processivity experiments. This bending estimate is consistent with both crystallographic and solution studies (40,41). The integrity of roadblock experiments critically relies on evidence of site specific occupancy.

Figure 4.2. Confirmation of site-specific occupancy of EcoRV ENase. **A**) Schematic of 3 experiments: i) A pre-incubated EcoRV, Ca^{2+}, and DNA complex undergoes cutting with the addition of Mg^{2+}. ii) The cutting proceeds in spite of the addition of excess DNA chase which is added simultaneously with the Mg^{2+}, suggesting that EcoRV remains bound at its site. Increasing enzyme shows no added cutting, displaying enzyme saturation. iii) The lack of a burst of product formation from EcoRV preincubated with chase DNA shows that the chase is capable of partitioning the enzyme, if it had left the site in ii. **B**) Demonstration of burst data for i-iii by electrophoresis for single site (276 mer), and **C**) for double site substrate (276 mer, 17 bp between EcoRV sites).The Chase DNA (35 bp) is 5µM.



We adopt a catalytic DNA cutting chase assay to confirm site specific binding by EcoRV (10). EcoRV displays burst kinetics, where the cleavage of DNA by a pre-incubated enzyme-DNA complex is faster than the subsequent catalytic cycle, where product release is the rate limiting step (42). The chase DNA is unlabeled and the DNA which is pre-incubated with EcoRV is labeled. The ability to cut the labeled DNA in the presence of chase is

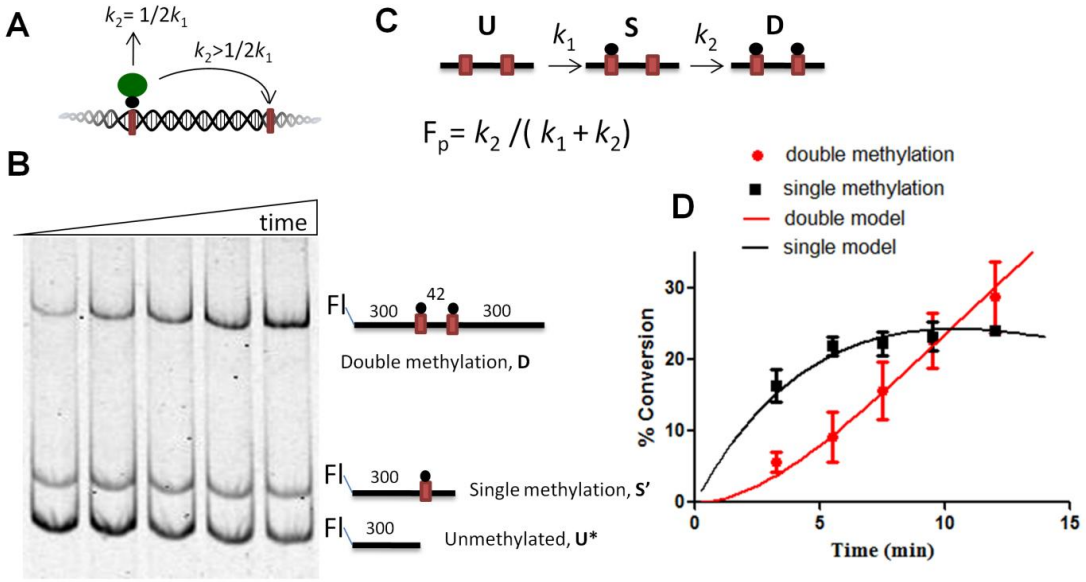
indicative of its specific binding as described in the following experiments (i-iii) (Figure 4.2A). In (i) enzyme is in slight excess and in the presence of Ca^{2+} is incubated with DNA, and cleavage is initiated with excess Mg^{2+}, resulting in total cleavage of the DNA (Figure 4.2B). The enzyme appears to exchange the divalent metal cofactors without perturbations of the DNA-protein complex, consistent with prior reports (41). In (ii) Mg^{2+} and excess chase DNA are simultaneously added to initiate the cutting reaction. This experiment addresses the possibility that EcoRV leaves the site but can rapidly return (which turns out not to be the case). Importantly, nearly all of the DNA is cut, showing the enzyme remains bound to the DNA (Figure 4.2A). The addition of more EcoRV does not result in further cutting, demonstrating EcoRV saturation (Figure 4.2B). In (iii) the chase and the labeled DNA are pre-incubated with EcoRV prior to the initiation of Mg^{2+}. The minimal burst of cut labeled DNA is indicative of the effectiveness of the chase DNA, where the cutting of the labeled DNA proceeds though the slow steady-state cycle (Figure 4.2B). These experiments (i-iii) were repeated with a two site substrate, again providing evidence of site specific occupancy (Figure 4.2C). Additional evidence for site specific binding is described below. Interestingly, the decrease in (ii) with respect to (i) is likely due to the delay in the cutting of the second strand (within one site) which may allow the capture of the enzyme by the chase DNA, as has been seen previously (10).

4.3.2 Processivity Assay

Our processivity assay involves excess DNA (400nM) with two Dam recognition sites, and low enzyme (7nM) to drive conditions where one enzyme or less binds to a single DNA molecule (Figure 4.3) (3). The reaction is initiated with the enzyme under multiple turnover conditions, and aliquoted time points are heat quenched, slow cooled, then subjected to a

DpnII digestion, which cleaves unmethylated Dam sites. PAGE analysis allows the determination of the relative amounts of unmethylated, singly methylated (S), and doubly methylated (D) DNA at each time point (Figure 4.3B,C). A minimal accumulation of singly methylated products corresponds to a highly processive reaction and a large accumulation of singly methylated products corresponds to a non-processive reaction.

Figure 4.3. Processivity Assay: **A**) Dam (green circle) immediately following the first methylation (black circle) at a GATC site (bold red lines). Processivity is based on the probability of a second methylation event following an initial one during a single binding event. **B**) The reaction includes 7nM Dam, 400nM DNA (symmetric substrate, fluorescein labeled), 30μM SAM (S-adenosyl methionine), 30 mM NaCl, 5mM $CaCl_2$, and buffer at 37°. Heat quenched samples were digested with DpnII endonuclease (cuts unmethylated GATC sites), and reaction products were separated by non-denaturing PAGE and imaged using a typhoon phosphoimager. **C**) The relative amount of each species of the reaction is fit to a sequential reaction mechanism model (Equations 1-3) to derive the rate constants k_1 and k_2, and ultimately the processivity value (F_p). **D**) Fit of data to model for substrate with 42 base pairs between sites.



The processivity data is quantified using a sequential methylation scheme, where k_1 is the first order rate constant from unmethylated to singly methylated DNA, and k_2 is the first order rate constant from singly to doubly methylated DNA (Eq. 1,2) (Figure 4.3C) (43).

Singly methylated (S) = $\frac{k_1}{(k_2-k_1)} \times (e^{-tk_1} - e^{-tk_2})$ (**Eq. 1**)

Doubly methylated (D) = $1+\frac{1}{(k_1-k_2)} \times (k_2 e^{-tk_1} - k_1 e^{-tk_2})$ (**Eq. 2**)

Processivity (F_p) (the relevant outcome of the analysis) relates the values of k_1 and k_2 (Eq. 3), which are ultimately derived from the gel densitometry as discussed above.

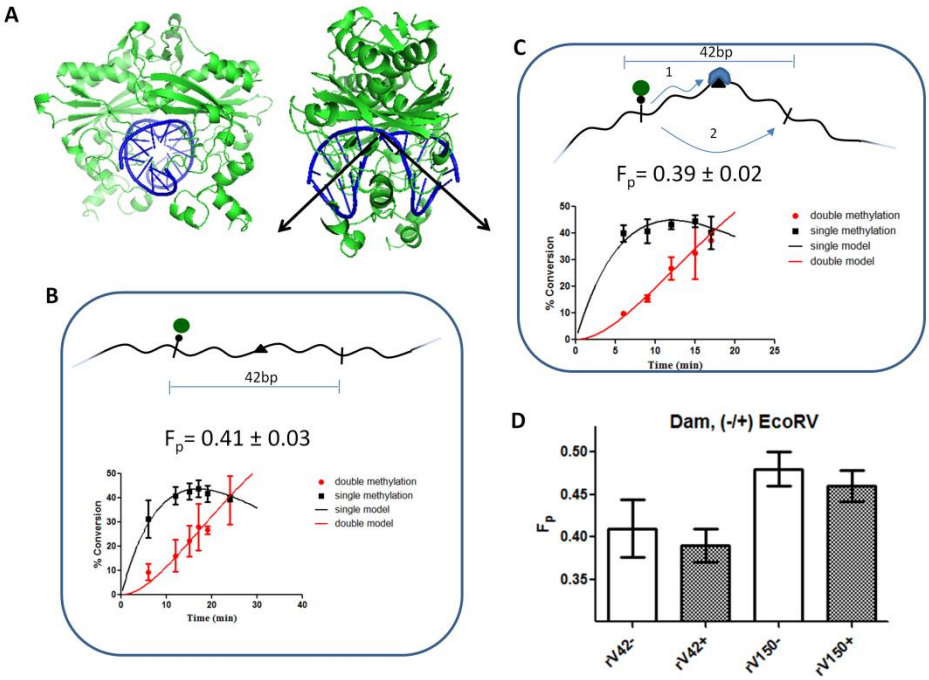
$F_p = \frac{k_2}{k_1+k_2}$ (**Eq 3**)

When $k_2 >> k_1$ processivity is 1.0, which is maximal. When $k_2 = k_1/2$, processivity is 1/3, which is minimal.

4.3.3 Dam's processivity is unaffected by an EcoRV roadblock

EcoRV binds DNA as a dimer, and the complex has ~ 5nm diameter (40,44). It almost completely encloses the DNA, and would therefore certainly sterically block an enzyme sliding along the DNA (Figure 4.4A) (40), as has been demonstrated previously with a similar noncatalytic EcoRI roadblock (36). Dam's processivity is minimally reduced by the EcoRV roadblock, as EcoRV's absence and presence results in small modulations in processivity for a substrate with 42 bp separating the two Dam sites (rV42), with the EcoRV site positioned in the middle of the two Dam sites (Figure 4.4B,C). The overall rate of the reaction is similar with and without roadblock (see below). Processivity is also minimally reduced using a substrate with 150 bp between the Dam sites and the EcoRV site in the middle (Figure 4.4D). This longer substrate (rV150) has a higher base-level (no roadblock) processivity, due to its greater DNA loop formation propensity (3).

Figure 4.4. Dam is processive in spite of an EcoRV roadblock. **A**) Structure of EcoRV dimer bound to cognate DNA (PDB 2RVE) (40) (left). Note the majority of the DNA is enclosed by this "saddle" dimer. EcoRV bends DNA ~50°, sharply between the T-A step (right). Arrows show the trajectory of the DNA from the bending locus. **B**) Dam's (green circle) processivity in the absence of the roadblock (the site is a black triangle) (rV42-) (276mer, 42 bp between Dam sites). **C**) Processivity in the presence of the EcoRV roadblock (blue shape) (rV42+) is relatively unchanged in comparison to **B**. The enzyme therefore likely does not travel along the helix using sliding or hopping (path 1). **D**) F_p data from **B** and **C** is plotted. The lack of effect of the roadblock is similar when the distance between the Dam sites is larger (rV150-/+) (384mer, 150 bp between sites). This suggests that Dam is unlikely to utilize sliding and hopping mechanisms, and instead favors an intersegmental hopping mechanism.



These processivity results suggest that Dam does not utilize sliding or short hops, and are provocative considering prior reports, where similar roadblocks vary from partially to completely inhibiting movement between sites (10,36). However, our data is consistent with previous results for Dam, which suggests its limited ability to track along the trajectory of the helix by sliding or hopping (3). A probable interpretation of this result invokes intersegmental hopping by DNA looping (see Introduction), which we predict would not be

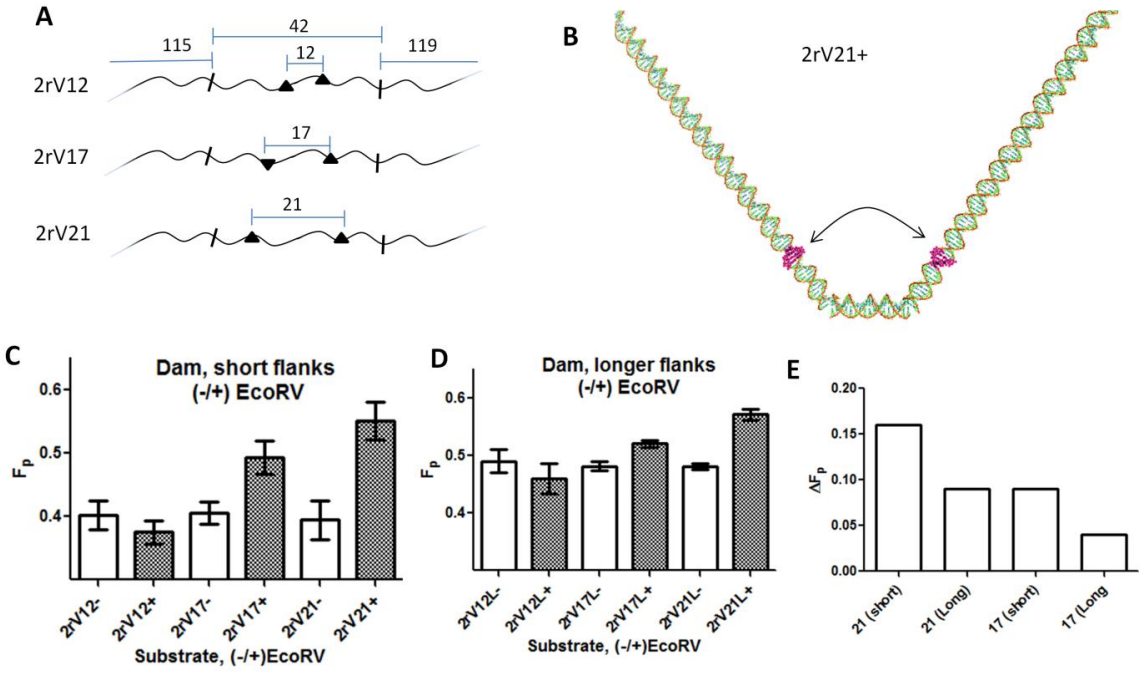
significantly disturbed by a roadblock. This involves the enzyme making a series of nonspecific intersegmental hops between DNA regions that mostly lack sites (Figure 4.1C), transitioning primarily on the DNA flanks, eventually resulting in processive catalysis.

The EcoRV roadblock may have pleiotropic effects, where the protein-induced bending of the DNA may enhance looping (and therefore processivity), but the roadblock may in other ways diminish processivity, for example, by sterically interacting with the traveling Dam enzyme. Our data seems consistent with an intersegmental hopping mechanism; however, experiments further modulating the ability of DNA to form loops (and enhance processivity) were carried out to provide a clearer mechanistic interpretation. Furthermore, minimal effects on Dam reaction rates by EcoRV suggest limited interactions between Dam and EcoRV (see below).

4.3.4 Two EcoRV roadblocks *increase* processivity by DNA looping

The lack of an effect on Dam's processivity with a single EcoRV roadblock motivated us to probe the potential processivity modulations from various configurations of two EcoRV roadblocks. We hypothesized two potential outcomes: 1) the processivity will decrease due to the roadblock(s) impeding hopping, as has been shown (10). Dam may have been able to hop over one centrally located roadblock (Figure 4.4D), but the sterics of two roadblocks positioned various distances from the Dam sites may block the enzyme and diminish processivity. 2) The EcoRV induced bending of DNA is combinatorial and will facilitate DNA looping; this may enhance intersegmental hopping and therefore increase processivity. The substrates with the two EcoRV sites are shown in Figure 4.5A, where the Dam sites are separated by 42 bp for each.

Figure 4.5. Dam processivity increases as a consequence of an EcoRV mediated enhancement in DNA looping. **A**) Schematic of the substrate series where the distance between Dam sites remains constant, yet the EcoRV roadblock sites (black triangles) are modulated. **B**) Schematic of 2rV21+ DNA using the 3-D Dart program (47,48). The EcoRV induced bending of the DNA likely enhances loop formation and subsequently Dam's ability to undergo intersegmental hopping. **C**) Processivity for Dam with and without EcoRV for the substrates in A, which here are 276 bp long. Processivity is most increased when DNA looping is optimal 2rV21+. **D**) Longer DNA flanks (642 bp total substrate size) enhance Dam's ability for intersegmental hopping in the absence of roadblocks. EcoRV roadblocks offer more modest increases in processivity due to the higher base-line processivity levels. **E**) ΔF_p from 5C (short flanks) and 5D (long flanks). Since ΔF_p decreases, the source of processivity enhancement is likely due to looping, and not by some other mechanism.



The phasing of DNA is 10.5 bp, and bends that are separated by such phasing have been shown to be additive and therefore can dramatically increase DNA loop formation in a variety of systems (45, 46). 2rV21+ includes this optimal phasing. Using the 3D-DART program (part of the HADDOCK system) (47,48), we construct what we imagine 2rV21+ looks like by manually introducing the 50° bends due to the EcoRV into the DNA structure (Figure 4.5B). The phasing of the substrate places the DNA flanking the EcoRV bends in the

same general space, allowing the GATC sites to be positioned close in space such that a combination of non-specific intersegmental hopping and possibly site specific intersegmental hopping can occur.

As expected, the processivity values for 2rV12-, 2rV17-, and 2rV21- are similar (Figure 4.5A,C). Consistent with hypothesis 2, Dam's processivity increases in the presence of the roadblocks. We also note that the differing processivity outputs based on modulations of where the EcoRV sites are is further evidence of site specific binding of the EcoRV roadblock(s). The most significant increase in processivity is seen with 2rV21+ which is consistent with the super-structured substrate in Figure 4.5B. The processivity data shows a clear trend with processivity increasing as the roadblocks approach the Dam sites. By orienting the Dam sites and the flanking DNA towards the center of the substrate, the ability of the protein to exchange between DNA flanks and Dam sites should be enhanced and cause an increase in processivity. To further show that the enhancement in processivity observed with this set of EcoRV substrates is not due to the proximity of the EcoRV protein to the Dam site, we designed the following set of experiments.

4.3.5 Modulating DNA flanks provides evidence for intersegmental hopping by EcoRV roadblocks

We sought to challenge the hypothesis that the enhancement in processivity in Figure 4.5 is due to DNA bending-mediated intersegmental hopping, and not by some other mechanism, involving, for example an interaction between the EcoRV roadblock and Dam, or allosteric modulation of the Dam site through the DNA upon EcoRV binding (49). We consider this in part because of the proximity of the EcoRV sites to the Dam sites. We chose to lengthen the DNA surrounding the Dam sites on the substrates from the previous section

(Figure 4.5A) that contain the two EcoRV roadblock sites separated by various amounts (Figure 4.5A). The lengthening of the flanking DNA increases processivity in the absence of roadblocks (Figure 4.5D) as the longer DNA flanks enhance DNA looping (as has been previously demonstrated (3)). We are interested in how the ΔF_p values (the F_p in the presence minus the F_p in the absence of roadblock) change between the current (Figure 4.5D, longer) and prior (Figure 4.5C, shorter) substrate set. We propose two possible outcomes: 1) If intersegmental hopping is involved, the ΔF_p with the longer substrates will be less. This is because the longer flanks enhance processivity, and the bending of the flanks towards each other may be redundant. 2) If another mechanism contributes to roadblock mediated enhancing of processivity, the ΔF_p will be the same regardless of the changes in flanking DNA.

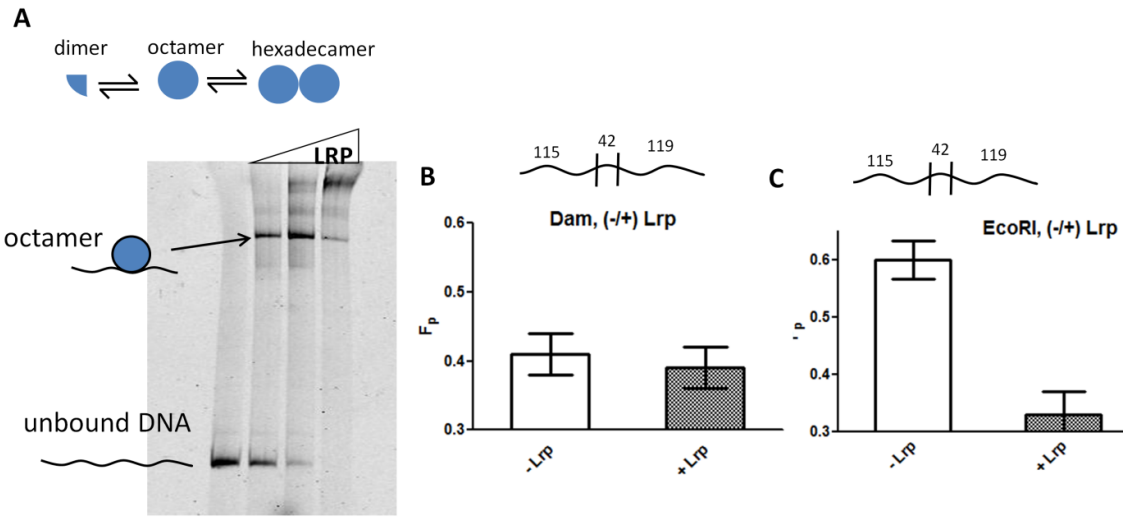
Our data supports hypothesis 1, where the ΔFp is less for the substrates with the longer flanks (Figure 4.5E). The inclusion of the roadblocks, with or without the flanks appears to maximize the enhancement in processivity as mediated by intersegmental hopping. For example, lengthening the flanks in Figure 4.5B is inconsequential based on similar F_p values for the 2rV21+ and 2rV21L+ (Figure 4.5C,D).

4.3.6 Nonspecific Lrp binding disturbs processivity for EcoRI, but not Dam

How do transient protein complexes differentially impact translocating proteins which rely on sliding versus intersegmental hopping? Here we rely on Lrp, a well-studied NAP. The Lrp-DNA complex moves our investigation towards a more biologically relevant situation. We confirmed conditions where Lrp, most likely as an octamer, binds DNA nonspecifically by electrophoresis (Figure 4.6A) (18). Similar tight banding (lack of smearing) of both nonspecific and specific DNA octamer complexes for Lrp has been

demonstrated (18). The lack of a smeared octamer band (Figure 4.6A) suggests that the Lrp is likely bound near the center of the DNA, as was demonstrated with the specific complex (18). If Lrp binds near the ends and in the center of the DNA, the banding would be much more smeared as different complexes have distinct gel mobilities.

Figure 4.6. Histone-like Lrp roadblock reveals a translocation mechanism dependent processivity response **A**) Lrp is capable of dynamic oligomerization, and can bind tightly to nonspecific DNA as an octamer. Gel shows stable nonspecific DNA-octamer complex. From left to right the lanes include 0, 0.5, 1 and 2 Lrp octamers per DNA. **B**) Dam's processivity is minimally affected by the Lrp roadblock. **C**) EcoRI ENase significantly utilizes sliding, and is nearly maximally processive on a substrate with 42 bp site separations (3). In the presence of Lrp, EcoRI's ability to undergo processive catalysis is essentially completely abolished. This demonstrates that Lrp can disrupt processivity.



Previous work has demonstrated that Dam's catalysis is relatively unaffected by non-specifically bound Lrp (50). Even though catalysis also involves the searching of the DNA for the Dam site, we sought to extend this previous result using our processivity analysis here. Dam's processivity is largely unaffected by the non-specifically bound Lrp roadblock for a substrate with 42 bp between the sites (Figure 4.6B). However, the transient nature of

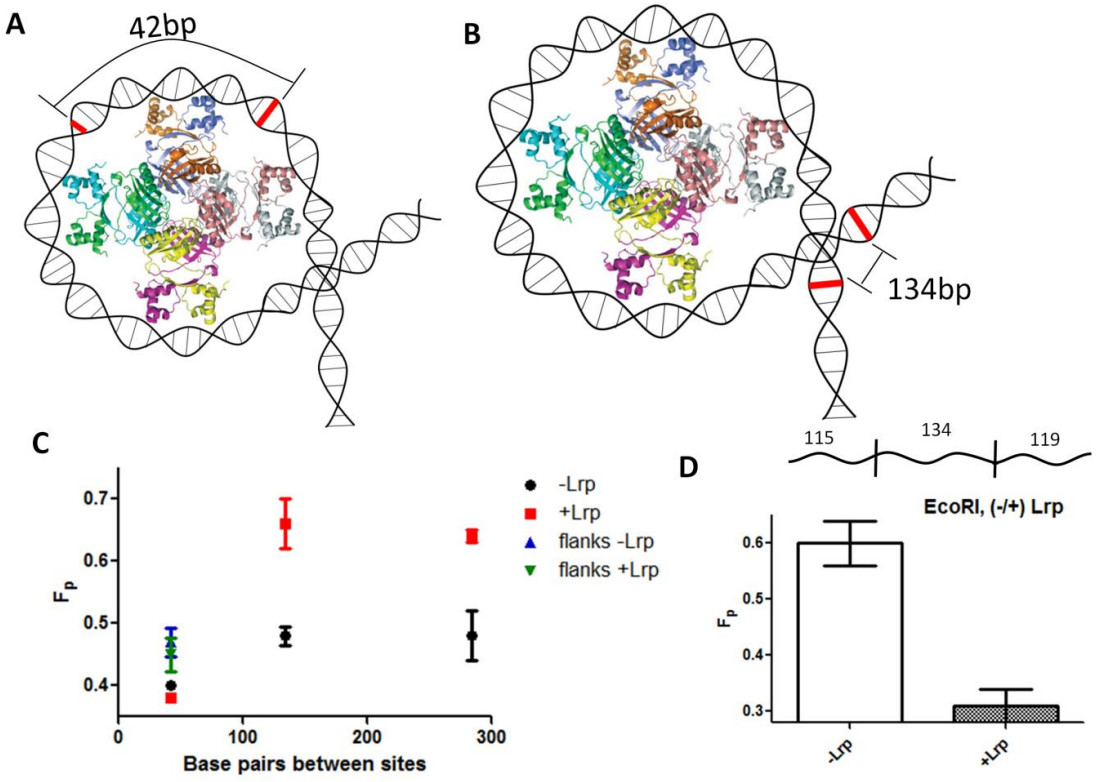
the Lrp-DNA nonspecific complex invokes the possibility that it is unable to block any type of facilitated diffusion process, which we address next.

In part as a control to confirm the binding of the Lrp and its potential as an effective roadblock, we carried about a similar experiment probing the EcoRI's processivity response to nonspecifically bound Lrp. Importantly, EcoRI's translocation mechanism is quite distinct from Dam's, where Dam's propensity for extensive sliding is unlikely (3). In contrast, several lines of evidence suggest that EcoRI primarily uses a sliding mechanism (3,8), where translocation proceeds with close association between the protein and the DNA (9). Therefore, we predict that EcoRI's processivity will be decreased by Lrp. We previously demonstrated that due to the cutting nature of endonucleases' catalytic activity, that their processivity range is 0.33-0.58 (see chapter 3). In the absence of the Lrp roadblock, EcoRI is nearly completely processive on a substrate with 42 bp site spacings (Figure 4.6C). Following the confirmation of the Lrp-DNA complex formation under the EcoRI reaction conditions by electrophoresis (similarly to Figure 4.6A, data not shown), we show that EcoRI's processivity is indeed dramatically decreased upon the addition of Lrp (Figure 4.6C). The processivity drop essentially spans the entire processivity range. This result confirms both the integrity of the non-specifically bound Lrp roadblock (demonstrating that it can disrupt protein movement) and that EcoRI's sliding shows the expected response to the large Lrp roadblock. This provides a proof-of-principle that proteins will respond differently to roadblocks based on their translocation mechanism.

4.3.7 Larger Dam intersite separations support Lrp based enhancements in processivity for Dam

A direct challenge of Lrp's ability to enhance intersegmental hopping by DNA looping relies on experiments using different spacing between the Dam sites. The crystal structure of *E. coli* Lrp is available (the DNA cannot be resolved in the structure). DNA has been modeled around it, resulting in a ~130 bp DNA loop surrounding the octamer (51). The hypothesized bending and looping of DNA by *E. coli* Lrp is consistent with the behavior of other proteins similar to Lrp (16,52,53, Discussion). Importantly, Lrp must dramatically manipulate the DNA, as ~130 bp DNA loops are nearly impossible to form in Lrp's absence. Therefore the two Dam sites separated by 42 bp are unlikely to be accessible to Dam and looped towards one another (Figure 4.7A). In contrast, when the sites are separated by 134 bp it becomes possible for the Dam sites to be looped proximally as a consequence of having the DNA looped around the Lrp octamer (Figure 4.7B). When separated by 284 bp the Dam sites again are unlikely to be imbedded within the Lrp-DNA complex, and instead are likely to be positioned proximally along the DNA flanks. Finally, EcoRI's processivity with Lrp when the EcoRI sites are separated by 134 bp is explored. The similar baseline (no roadblock) processivity between 42 and 134 bp spacings is consistent with prior results demonstrating EcoRI's extensive use of sliding (see chapter 3) (Figure 4.7D). The significant drop in processivity to essentially the lower limit of processivity is further suggestive of the inability of EcoRI to utilize intersegmental hopping (Figure 4.7D), and its extensive use of sliding. This result further legitimizes the interpretation that Dam's site spacing specific Lrp dependence is due to intersegmental hopping.

Figure 4.7. Lrp enhances Dam processivity only when it is able to loop Dam sites into proximity. **A**) The crystal structure of *E. coli* Lrp (2GQQ) with DNA modeled in. ~130 bp of DNA are wrapped around Lrp (51). The Dam sites (red) are ~42 bp apart. **B**) The same Lrp complex (from A) with DNA where the distance between Dam sites is ~134 bp. When the sites are far enough separated on DNA, it becomes possible that Lrp loops the Dam sites proximally. **C**) When the Dam sites are separated by 134 bp and 284 bp, Lrp's looping increases Dam's processivity. For two substrates when the Dam sites are 42 bp apart processivity is not enhanced by Lrp: 1) replot of data from Figure 4.6B (red/black, no error); 2) the substrate called "flanks" (blue/green), which refers to a substrate with 300 bp flanks on either side of the Dam sites. **D**) EcoRI's processivity on a substrate with 134 bp separating the sites is significantly decreased in the presence of Lrp, consistent with EcoRI's inability to utilize intersegmental hopping.

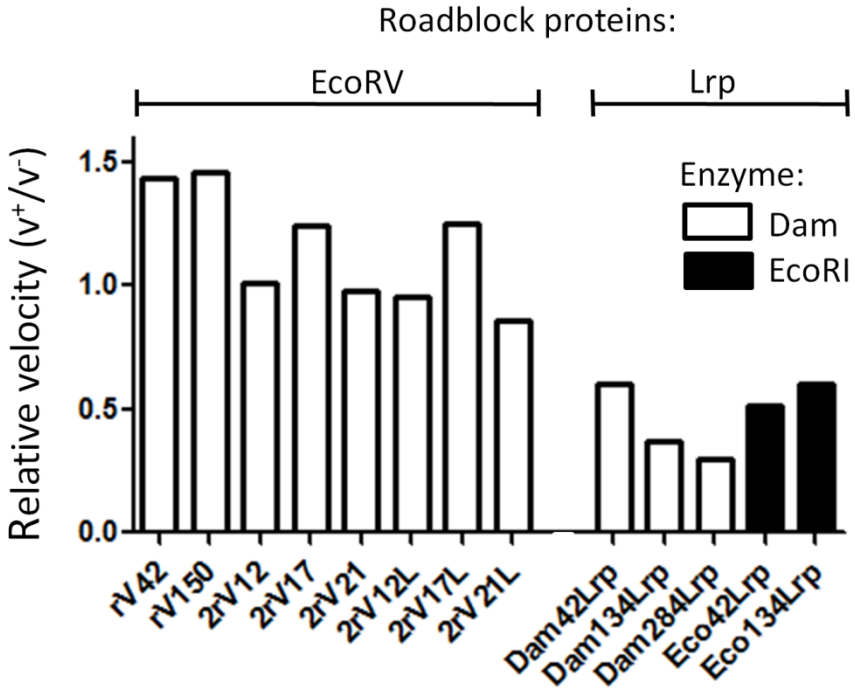


4.3.8 Roadblock mediated velocity modulations

When either 134 or 284 bp separate the sites, the processivity is increased with the presence of Lrp (Figure 4.7C). We stress the novelty of this result, as to the best of our knowledge this is the first demonstration that histone-like proteins can *enhance* facilitated

diffusion processes. We believe this is due to the ability of the Lrp to loop the segments containing the Dam sites into proximity (which is unavailable in the 42 bp Dam site separation in Figure 4,7A,C). As a control we tested the effect on processivity for the 42 bp separation where the flanking DNA segments are increased. Consistent with the previous result (Figure 4.6B), the minimal decrease in processivity due to the inability of the Dam site containing DNA segments to loop is further suggestive that the spacings between the sites is the critical parameter, not simply the length of the DNA (Figure 4.7C).

Figure 4.8. Reaction rates are predictably modulated by roadblock type. Velocity (k_1 + k_2) with roadblock is divided by velocity in the absence of roadblock from processivity experiments (Figures 4.4-4.7). EcoRV roadblock(s) data (Figures 4.4 and 4.5) is on the left, Lrp roadblock data (Figures 4.6 and 4.7) is on the right. Dam processivity is clear bars, EcoRI processivity is filled bars. EcoRV does not decrease velocity while Lrp does, consistent with Lrp's significant alteration of DNA structure.



Our interest in processivity is critical to deduce translocation properties. However, the relative velocities (product formation over time) in the presence (and absence) of roadblocks may provide information about our experimental systems. A single EcoRV roadblock (rV42 and rV150) causes a slight increase in velocity (Figure 4.8 (from figure 4.4)). In general, two EcoRV roadblocks seem to have little effect on the reaction rates, and interestingly the velocity is slightly increased when the EcoRV sites are separated by 17bp (Figure 4.8 (from Figure 4.5)). These results are consistent with Dam minimally interacting with EcoRV, which is likely binding tightly on its recognition site (representing a small fraction of the total DNA size). Importantly, if the EcoRV roadblocks were interfering with Dam when in close proximity (2rV21 and 2rV21L: Figure 4.5), the velocity would likely decrease, which is not the case; for example, the minimal velocity differential is similar between 2rV12 and 2rV21, where in the former the EcoRV sites are much farther from the Dam sites than in the later. In contrast, Lrp lowers the rates ~two fold for both Dam and EcoRI reactions (Figure 4.8 (from Figures 4.6 and 4.7)). The decrease in velocity is predictably due to Lrp's large size and extensive interactions with the DNA.

4.4 Discussion

The majority of facilitated diffusion studies are by necessity constrained to conditions significantly removed from any biological context. Considering how the movement of proteins along DNA is impacted by other proteins is an important step towards reconciling the *in vivo* legitimacy of the detailed mechanistic insights gained from facilitated diffusion studies. Furthermore, well established sliding and hopping mechanisms have been studied with DNA molecules that are typically less than 100-200 bp, while site finding in the cell likely involves longer movements. Also, only a minority of DNA binding proteins'

facilitated diffusion properties have been studied. Intersegmental hopping represents an intuitive example of previously ambiguously described 3-D movements that account for protein translocation on the order of thousands of bp. However, while growing evidence suggests that other proteins (besides Dam) likely use intersegmental hopping (28-31), the consideration of this mechanism in the presence of roadblocks remains experimentally unexplored. In general, the extent of large, histone-like proteins, whose abundance in both bacteria and eukaryotes is massive, is only now being investigated in its relation to facilitated diffusion (22).

HU (heat-unstable nucleoid protein), considered a prokaryotic histone homolog, is one of the bacterial NAPs that cover the *E. coli* genome (15). HU's ~ 50,000 molecules in the cell support a separation of, on average, ~350 bp between HU oligamers (15). HU induces DNA loop formation, where ~100mers cyclize in the presence of nonspecifically bound HU. This involves numerous contacts between the protein and the DNA that significantly alter the structure of DNA to achieve such looping, similarly to what has been seen with *E. coli* Lrp (and related proteins) (16,51-53). These multiple structural and biochemical examples provide the basis for the schematics in Figure 4.7A,B. Interestingly, the 11.5 nm diameter of the related MtbLrp is similar to the 11 nm nucleosome diameter (52), suggesting that experiments with Lrp can be extrapolated to understanding movement in the eukaryotic cell.

Our results with Dam and Lrp show that Lrp's ability to loop DNA provides an enhancement in processivity (Figure 4.7). Importantly, the increase only occurs when the Dam sites are sufficiently separated to be looped by Lrp, consistent with an intersegmental hopping mechanism (Figure 4.7). Furthermore, the two site EcoRV (Figure 4.5) and Lrp results (Figure 4.6 and 4.7) both demonstrate a context-specific roadblock-enhancement of

Dam's processive catalysis, which taken together provides added legitimacy for these provocative results. Intersegmental hopping is an elegant way for proteins to navigate along DNA in the presence of roadblocks which may appear to confound the site finding process. Statistically, Dam sites are separated by ~250 bp in the *E. coli* genome, further suggesting the biological relevance of intersegmental hopping being enhanced by NAPs (Figure 4.7). In addition, known and putative Dam sites likely to undergo epigenetic gene regulation are typically clustered (<50 bp separation) (37), and are hypothesized to undergo *unprocessive* catalysis to maintain their epigenetic state (54). Lrp's inability to enhance the processivity of closely spaced Dam sites (Figures 4.6 and 4.7) is consistent with this.

Many theoretical studies have considered proteins' compaction of DNA as a mode of enhancing site finding. For example, Li and coworkers suggested that the amount of DNA binding proteins in an *E. coli* cell requires that proteins utilize a looping-based translocation mechanism (21). These predictions are consistent with our experimental data and the notion that "roadblocks" can help with facilitated diffusion processes (23, 24). Furthermore, these studies suggest that sliding or hopping processes that involve movement along the trajectory of the helix (Figure 4.1A) are unlikely to contribute to site finding *in vivo*, and that many DNA binding proteins will likely use looping-based translocations that intersegmental hopping accommodates. Our results speak directly to the uncertainty in the literature regarding the potential consequences of DNA-associated proteins on facilitated diffusion properties (Introduction).

As a control to our Dam-Lrp studies, we show EcoRI cannot bypass the Lrp roadblock (with either 42 or 134 bp separations), and consequently does not utilize intersegmental hopping as a significant contribution to its site finding process, consistent with prior reports

(3,8). Our results show that these Lrp experiments are able to dramatically differentiate between translocation mechanisms. Dam and EcoRI rely on distinctive blends of translocation mechanisms which is further demonstrated here, showing the value of these roadblocks as a tool for probing the details of translocation mechanisms for a variety of other proteins of interest.

Several techniques and approaches have been used to elucidate the details concerning translocation mechanisms, including single molecule studies (55). The most commonly used strategy, however, involves predictable modulations in processivity values with respect to changes in the intersite spacings between target sites (Introduction). Yet many enzymes have been shown to utilize a combination of sliding and hopping, and segregating the extent of each can be quantitatively and mechanistically unsatisfying (considering intersegmental hopping further complicates this). Clever modifications to such processivity assays can differentiate between translocation mechanisms (10,31,56,57). The systematic variation of site to site distances, combined with a blocking protein (Figure 4.4) may be useful for studying other enzymes. For example, an enzyme's (one that uses a combination of sliding and hopping (9-11)) processivity may be decreased when the roadblock is included between closely spaced sites, but will have less of an effect when the sites are farther apart. Such a trend can be informative to estimate the relative bp range for a hop.

Dam's overwhelming use of intersegmental hopping is reconciled by its low copy number, need to transverse large regions of the genome (3), and minimal need to make short movements along DNA (Introduction). Dissimilarly, EcoRI is involved in a restriction modification system, where it cleaves invading foreign nucleic acids, such as phage DNA. Incoming phage DNA is much less likely than genomic DNA to be associated with NAP's,

in part to ensure the transcription of phage genes (58,59). This allows EcoRI to slide to find its sites on phage DNA. EcoRI's inability to maneuver among histone-like roadblocks may also contribute to the prevention of cutting genomic DNA. However, other proteins (than Dam), such as many transcription factors, with low copy numbers (60) and rare binding sites (12), may also utilize intersegmental hopping. These roadblock types of experiments may help to reveal this.

4.5 Materials and Methods

Processivity assays: (Dam) Reactions were done in MRB (methylation reaction buffer: 100mM Tris H-Cl: pH8.0, 10 mM EDTA (ethylenediaminetetraactetic acid), 1mM DTT (dithiothreitol), 0.2mg/mL BSA (bovine serum albumin)) with 400nM DNA, 5mM $CaCl_2$ (only for EcoRV-/+ reactions), 7nM Dam, and 30 μM AdoMet (S-adenosyl methionine, SAM) at 37°. Roadblocks (where applicable) were pre-incubated for 20 minutes prior to the initiation of the reaction by Dam. Single site EcoRV roadblock experiments (Figure 4.4) included 600nM EcoRV dimers. Two site EcoRV roadblock experiments (Figure 4.5) included 800nM EcoRV dimers. Lrp roadblock experiments (Figure 4.6 and 4.7) included 400nM Lrp octamers. Dam was diluted in Protein Dilution Buffer (20mM potassium phosphate, 200mM NaCl, 0.2mM EDTA 0.2 mg/mL BSA 2mM DTT and 10% glycerol). Aliquots of the reaction mixture were heat inactivated in 75° water, which has previously been shown to be effective (3,37). After slow cooling, NEB DpnII Buffer and DpnII enzyme was added and the mixture was incubated at 37° for 2 hours. The reaction products were analyzed using PAGE (7.5% 29:1 acrylamide:bisacrylamide).

(EcoRI ENase) Reactions were done in ERB (endonucleases reaction buffer: 20 mM Tris: pH 7.4, 5mM $MgCl_2$, 0.2 mM EDTA, .05 mg/ml BSA) with 400nM DNA, 65mM

NaCl, and 1 nM EcoRI ENase at 37°. Lrp roadblocks (400nM octamers, where applicable) were pre-incubated for 20 minutes prior to the initiation of the reaction by EcoRI. Reactions were quenched in 50mM EDTA then run on a PAGE gel as above.

Gel shifts: Lrp was added to both MRB and ERB (see above) with 400nM DNA for 20 minutes. The mixture was quenched with an equal volume of 50% glycerol and immediately run on a gel.

EcoRV binding Chase assay: 400 nM DNA (rV42 in Figure 4.2B, 2rV17 in Figure 4.2C) was added into MRB with 5mM $CaCl_2$ at 4°. EcoRV was added at indicated concentrations. $MgCl_2$ was added to initiate reactions to a total concentration of 10mM. In ii, $MgCl_2$ was mixed with unlabeled chase DNA which was simultaneously added generating a total of concentration of 5μM chase DNA. In iii, the chase DNA (total concentration of 5μM) was added prior to the addition of the EcoRV.

Enzyme expression and purification: Dam was purified exactly as before (37). EcoRV was kindly provides for by Dr. John Perona, and was purified as described (41). Lrp was purified as previously described (18). EcoRI ENase was purchased from NEB.

DNA Substrates: All restriction endonucleases for cloning were purchased from NEB. All synthetic DNA substrates and primers were purchased from IDT and were re-suspended in TE buffer (10mM Tris, pH 7.5, 1mM EDTA). When necessary, DNA was annealed with its reverse complement in a 1:1 mixture for 5 minutes at 95 degrees, and allowed to slow cool to room temperature (~5 hours) to generate oligonucleotides. Annealing was verified by PAGE. The production of substrates used the following general strategy: A synthetic oligonucleotide with two GATC sites (below), and two restriction sites between the GATC sites was cloned into plasmid pBR322 (NEB), making a new plasmid, pBR322GATC. The

same strategy was used to generate the EcoRI Enase substrates making another plasmid, pBR322GAATTC. Spacers (PCR derived and synthetic oligonucleotides) were cloned into pBR322GATC and pBR322GAATC, generating several different plasmids with different distances between the GATC or GAATTC sites and different amounts and locations of EcoRV sites for roadblocks. These plasmids were PCR amplified with different primers to adjust the spacings from the GATC or GAATTC sites to the ends of the DNA. PCR amplicons were purified using the Agilent PCR clean-up kit.

The following oligonucleotide (described above) was cloned into the plasmid pBR322 at the EcoRI and HindIII sites, 5'-aattcggtgatctcgtcga*cccggg*agctggtagtatgc*ccatgg*ttcgatcggatgcca-3'; 5'-agcttggcatccgatcgaa*ccatggg*catactaccagct*cccggg*tcgacgagatcaccg-3', making, pBR322GATC. The following oligonucleotide was cloned into the plasmid pBR322 also at the EcoRI and HindIII sites, 5'-agcttggcattgaattctaccatgggcatactaccagctcccgggtcggctcgatcaccg-3' ; 5'-aattcggtgatcgagccgacccgggagctggtagtatgcccatggtagaattcaatgcca-3', making pBR322GAATTC. The cloned, synthetic oligonucleotides (above) used to make pBR322GATC and pBR322GAATTC had additional cloning sites within each: XmaI and NcoI (italicized). These sites were used to insert PCR purified spacers or oligonucleotides (listed below) between the two (Dam or EcoRI) sites (underlined).

Annealed oligonucleotide spacers cloned into pBR322GATC at XmaI and NcoI sites:

rV42 (Figures: 4.2B,4.3,4.4B-D,4.6A-B,4.7C,4.8)

Top: 5'-ccgggtacaacttggatatcggttacacgcc-3'

Bottom: 5'-catgggcgtgtaaccgatatccaagttgtac-3'

2rV12 (Figures: 4.5A,4.5C-E,4.8)

Top: 5’- ccgggtacgatatcaccatagatatctggac

Bottom: 5’catggtccagatatctatggtgatatcgtac

2rV17 (Figures: 4.2C,4.5A,4.5C-E,4.8)

Top: 5’-ccgggtgatatctactcaggttagatatcac 3’

Bottom: 5’- catggtgatatctaacctgagtagatatcac

2rV21 (Figures: 4.5A-E,4.8)

Top: 5’-ccgggatatctcgactcagagtctagatatc-3’

Bottom: 5’-catggatatctagactctgagtcgagatatc-3’

PCR spacers cloned into pBR322GATC at XmaI and NcoI sites. Upon PCR amplification, the following spacers were digested with XmaI and NcoI to generate overhangs. The spacers were generated by PCR from the plasmid pBR322.

rV150mer (Figures: 4.4D,4.8)

Forward: 5’-attcccgggtcatcctcggcaccgtcacc-3’

Reverse: 5’-taatccatggtccgagaacgggtgcgcata-3’

Dam134Lrp and Eco134Lrp (cloned into pBR322GAATTC): (Figures: 4.7C-D,4.8)

Forward: 5’- att*cccggg*ggctaccctgtggaacacct - 3:

Reverse: 5’- taatccatggtactggaacgttgtgagggt-3’

Dam284Lrp: (Figures: 4.7C,4.8)

Forward: 5’- att*cccggg*ggctaccctgtggaacacct - 3:

Reverse: 5’- taatccatggtgataaagcgggccatgtta-3’

Note: Eco42Lrp: (Figures: 4.7C,4.8) used pBR322GAATTC without the insert of a spacer.

Amplicons with 115/119 base pairs flanking GATC sites were amplified from each plasmid using primers: (forward) 5'- gggttccgcgcacatttccc-3' and (reverse) 5'-(Fl) ccagggtgacggtgccgagg-3'(From Figures (4.2,4.4,4.5C,4.6,4.7C,4.8).

Amplicons with 300 base pairs flanking GATC sites were amplified from each plasmid using primers: (forward) 5'- gcatcttttactttcaccagcg-3', and (reverse) 5'-(Fl)ggctccaagtagcgaagcgagc-3'. (From Figures 4.3,4.5D,4.7C,4.8)

rV chase DNA (Figure 4.2):

Forward:5'-Ccgggtacaacttggatatcggttacacgcccatg

Reverse:5'-catgggcgtgtaaccgatatccaagttgtacccgg

Data analysis: The processivity values were derived from a least squares fitting using Microsoft excel. Gels were scanned on a Typhoon Phosphoimager (GE). DNA modeling was done using 3D-DART (47,48).

4.6 References

1. Halford, S. E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans*. **37**, 343-348

2. Halford, S. E. & Marko J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res*. **32**, 3040-3052

3. Pollak, A. J., Chin, A. T., Brown, F. L. H. & Reich, N.O. (2014) DNA looping provides for "intersegmental hopping" by proteins: a mechanism for long-range site localization. *J. Mol. Biol.* DOI: 10.1016/j.jmb.2014.08.002

4. Horton, J. R., Liebert, K., Hattman, S., Jeltsch, A. & Cheng, X. (2005). Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. *Cell* **121,** 349-361

5. Vuzman, D., Azia, A. & Levy, Y. (2010) Searching DNA via a "monkey bar" mechanism: The significance of disordered tails. *J. Mol. Biol.* **396,** 674-684

6. Doucleft, M. & Clore, G. M. (2008) Global jumping and domain-specific intersegmental transfer between DNA cognate site of the same multidomain transcription factore Oct-1. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13871-13876

7. Terry, B. J., Jack, W. E. & Modrich,P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Non-specific interactions in EcoRI catalysis. *J. Biol. Chem.* **260**, 13130-13137

8. Rau, D. C. & Sidorova, N. Y. (2010) Diffusion of the restriction nuclease EcoRI along DNA. *J. Mol. Biol.* **395**, 408-416

9. Stanford, N. P., Szczelkun, M. D., Marko, J. F. & Halford,S.E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* **19**, 6546-6557

10. Hedglin, M. & O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. *ACS Chem. Biol.* **5**, 427-436

11. Porecha, R. H. & Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10791-10796

12. Mendoza-Vargas, A. et al. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One* **4**, e7526.

13. Casadesús, J., & Low,D. (2006) Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* **70**, 830-856

14. Riggs, A. D., Bourgeois, S. & Cohn, M. (1970) The lac represser-operator interaction: III. Kinetic studies. *J. Mol. Biol.* **53**, 401-417

15. Azam, T. A., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J. Bacterial.* **181**, 6361-6370

16. Luijsterburg, M. S., Noom, M. C., Wuite, G. J., & Dame, R. T. (2006). The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.* **156**, 262-272

17. Dillon, S. C. & Dorman, C. J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**, 185-195

18. Peterson, S. N., Dahlquist, F. W. & Reich, N. O. (2007)The role of high affinity non-specific DNA binding by Lrp in transcriptional regulation and DNA organization. *J. Mol. Biol.* **369**, 1307-1317

19. Flyvbjerg, H., Keatch, S. A. & Dryden, D. T. (2006). Strong physical constraints on sequence-specific target location by proteins on DNA molecules. *Nucleic Acids Res.* **34,** 2550-2557

20. Koslover, E. F., Díaz de La Rosa, M. A. & Spakowitz, A. J. (2011). Theoretical and computational modeling of target-site search kinetics *in vitro* and *in vivo*. *Biophys. J.* **101**, 856-865

21. Li, G. W., Berg, O. G. & Elf, J. (2009) Effect of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **5**, 294-297

22. Gorman, J., Plys, A. J., Visnapuu, M. L., Alani, E. & Greene, E. C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* **17**, 932-938

23. Marcovitz, A. & Levy, Y. (2013) Obstacles may facilitate and direct dna search by proteins. *Biophys. J.* **104**, 2042-2050.

24. Benichou, O., Chevalier, C., Meyer, B., & Voituriez, R. (2011). Facilitated diffusion of proteins on chromatin. *Phys. Rev. Lett.* **106**, 038102

25. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. (1997) Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6185-6190

26. Shore, D., Langowski, J. & Baldwin, R. L. (1981) DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4833-4837

27. Ringrose, L. & Chabanis, S. (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO J.* **18**, 6630-6641

28. Gowers, D. M. & Halford, S. E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO. J.* **22**, 1410-1418

29. van den Broek, B., Lomholt, M. A., Kalisch, S. M. J., Metzler, R. & Wuite, G. J. L. (2008) How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1578-15742

30. Sidorova, N. Y., Scott, T. & Rau, D. C. (2013) DNA concentration dependent dissociation of EcoRI: direct transfer or reaction during hopping. *Biophys. J.* **104**, 1296-13039

31. Hedglin, M., Zhang, Y. & O'Brien, P. J. (2013) Isololating contributions from intersegmental transfer to DNA searching by Alkyladenine DNA Glycosylase. *J. Biol. Chem.* **288**, 24550-24559

32. Wentzell, L. M. & Halford, S. E. (1998) DNA looping by the Sfi I restriction endonuclease. *J. Mol. Biol.* **281**, 433-444

33. Mashhoon, N., Carroll, M., Pruss, C., Eberhard, J., Ishikawa, S., Estabrook, R. A. & Reich, N.O. (2004). Functional characterization of *Escherichia coli* DNA adenine methyltransferase, a novel target for antibiotics. *J. Biol. Chem.* **279**, 52075-52081

34. Wion, D. & Casadesus, J. (2006) N^6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* **4**, 183-192

35. Kunkel, T. A. & Erie, D. A. (2005) DNA mismatch repair. *Annu. Rev. Biochem.* **74,** 681-710

36. Pluciennik, A. & Modrich, P. (2007) Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12709-12713

37. Pollak, A. J. & Reich, N. O. (2012) Proximal recognition sites facilitate intrasite hopping by DNA adenine methyltransferase: mechanistic exploration of epigenetic gene regulation. *J. Biol. Chem.* **287**, 22873-22881

38. Jeltsch, A., Alves, J., Wolfes, H., Maass, G. & Pingoud, A. (1994) Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA. *Biochemistry* **33**, 10215-10219.

39. Martin, A. M., Horton, N. C., Lusetti, S., Reich, N. O. & Perona, J. J. (1999). Divalent metal dependence of site-specific DNA binding by Eco RV endonuclease. *Biochemistry* **38**, 8430-8439

40. Winkler, *et al.* (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781-1795

41. Hiller, D. A., Fogg, J. M., Martin, A. M., Beechem, J. M., Reich, N. O. & Perona, J. J. (2003). Simultaneous DNA binding and bending by EcoRV endonuclease observed by real-time fluorescence. *Biochemistry* **42**, 14375-14385

42. Baldwin, G. S., Sessions, R. B., Erskine, S. G. & Halford, S. E. (1999) DNA cleavage by the EcoRV restriction endonucleases: roles of divalent metal ion in specificity and catalysis. *J. Mol. Biol.* **288**, 87-103

43. Fersht, A. (1998) *Structure and Mechanism in Protein Science*, W. H. Freeman, New York.

44. Horton, N. C., Newberry, K. J. & Perona, J. J. (1998) Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13489-13494

45. Zinkel, S. S. & Crothers, D. M. (1987) DNA bend direction by phase sensitive detection. *Nature* **328**, 178-181

46. Kahn, J. D. & Crothers, D. M. (1992) Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6343-6347

47. van Dijk, M. & Bonvin, A. M. (2009). 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.* **37, (Web Server Issue):W235-W239** doi:10.1093/nar/gkp287

48. De Vries, S. J., van Dijk, M. & Bonvin, A. M. J. J. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883-897.

49. Kim, S, *et al.* (2010) Probing allostery through DNA. *Science* **339**, 816-819

50. Peterson, S. N. & Reich, N. O. (2008) Competitive Lrp and dam assembly at the pap regulatory region: implications for mechanisms of epigenetic regulation. *J. Mol. Biol.* **383**, 92-105

51. de los Rios, S. & Perona, J. J. (2007). Structure of the *Escherichia coli* Leucine-responsive regulatory protein Lrp reveals a novel octameric assembly. *J. Mol. Biol.* **366**, 1589-1602

52. Shrivastava, T. & Ramachandran, R. (2007) Mechanistic insights from the crystal structures of a feast/famine regulatory protein from *Mycobacterium tuberculosis* H37Rv. *Nucleic Acids Res.* **35**, 7324-7335

53. Tapias, A., López, G. & Ayora, S. (2000). *Bacillus subtilis* LrpC is a sequence-independent DNA-binding and DNA-bending protein which bridges DNA. *Nucleic Acids Res.* **28**, 552-559

54. Peterson, S. N. & Norbert O. Reich. (2006) GATC flanking sequences regulate dam activity: evidence for how dam specificity may influence pap expression. *J. Mol. Biol.* **355**, 459-472

55. Gorman, J. & Greene, E. C. (2008) Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.***15**, 768-774

56. Gowers, D. M., Wilson, G, G. & Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15883-15888

57. Schonhoft, J. D. & Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA. *Nat. Chem. Biol.* **8**, 205-210

58. Dorman, C. J. (2004) H-NS: a universal regulator for a dynamic genome. *Nat. Rev. Microbiol.***2**, 391-400

59. Liu, Q, & Richardson, C. C. (1993) Gene 5.5 protein of bacteriophage T7 inhibits the nucleoid protein H-NS of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1761-1765

60. Ghaemmaghami, S. *et al.* (2003) Global analysis of protein expression in yeast. *Nature* **425**, 737-741

V. Proximal recognition sites facilitate intrasite hopping by DNA adenine methyltransferase: a mechanistic exploration of epigenetic gene regulation

Parts or sections of this chapter were taken with permission from: Pollak, A. J. & Reich, N. O. (2012) Proximal recognition sites facilitate intrasite hopping by DNA adenine methyltransferase: mechanistic exploration of epigenetic gene regulation. *J. Biol. Chem.* **287**, 22873-22881

5.1 Abstract

The methylation of adenine in palindromic 5'-GATC-3' sites by *E. coli* Dam supports diverse roles, including the essential regulation of virulence genes in several human pathogens. As a result of a unique hopping mechanism, Dam methylates both strands of the same site prior to fully dissociating from the DNA, a process referred to as intrasite processivity. The application of a DpnI restriction endonuclease-based assay allowed the direct interrogation of this mechanism with a variety of DNA substrates. Intrasite processivity is disrupted when the DNA flanking a single GATC site is longer than 400 bp on either side. Interestingly, the introduction of a second GATC site within this flanking DNA reinstates intrasite methylation of both sites. Our results show that intrasite methylation occurs only when GATC sites are clustered, as is found in gene segments both known and postulated to undergo *in vivo* epigenetic regulation by Dam methylation. We propose a model for intrasite methylation in which Dam bound to flanking DNA is an obligate intermediate. Our results provide insights into how intrasite processivity, which appears to be context dependent, may contribute to the diverse biological roles that are

carried out by Dam.

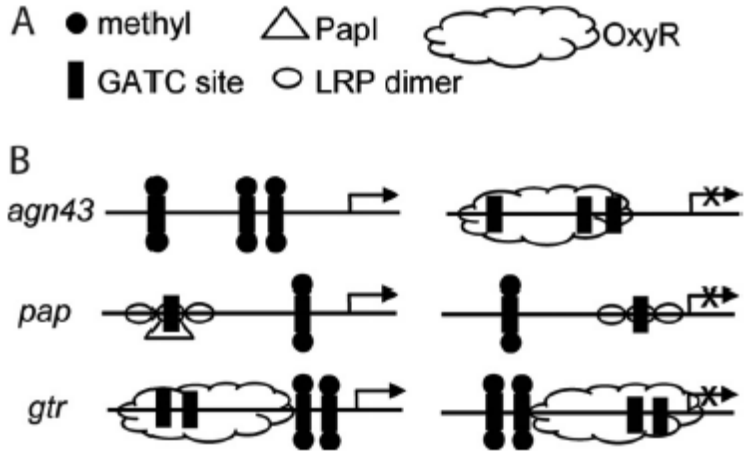
5.2 Introduction

The monomeric *E. coli* Dam methylates adenines at the N^6 position of palindromic 5'-GATC-3' sites (1). Unlike the majority of bacterial DNA methyltransferases, Dam lacks a cognate endonuclease. Dam is involved in the mismatch repair system (2), chromosome replication (3), nucleiod structure determination (4), and gene regulation (5, 6). Known and putative *dam* genes are co-conserved with MutH, a key protein for mismatch repair (7). More than 99% of GATC sites in the *E. coli* genome transition between hemi- and fully-methylated states (2). Hemi-methylated corresponds to only one of the two available adenines in the palidromic site being methylated. Immediately after replication, only the parental strand is methylated, which guides the mismatch repair system; these hemi-methylated GATC sites are Dam's substrate for mismatch repair. The newly synthesized daughter strand is initially unmodified, and the errors made in its synthesis resulting in mismatched bases are corrected by the mismatch repair system. The MutS recognizes the mismatched base pair. MutL and MutH associate with MutS, forming a ternary complex (2,3). The methylated adenine guides MutH to cut the DNA 5' to the strand of DNA including the methylated GATC site. The UvrD helicase then degrades the daughter strand well past the mismatched site, and DNA polymerase III re synthesizes the daughter DNA.

While nearly all of the 20,000 GATC sites in *E.coli* are involved in mismatch repair, ~0.1% of these are excluded from this process (8), and can be heritably unmethylated. However, these sites can be methylated upon differing environmental conditions (9), suggesting that they may be involved in gene regulation. The methylation state of a subset of these GATC sites epigenetically regulates gene transcription (Figure 5.1) such as the *pap*

(10), *gtr* (5), and *agn43* (11) promoters, where GATC sites switch between the unmethylated to the fully methylated states.

Figure 5.1. Known examples of methylation state dependent epigenetically regulated operons. (A) Legend. (B) Operons in the "on" state (left). Operons in the "off" state (right).



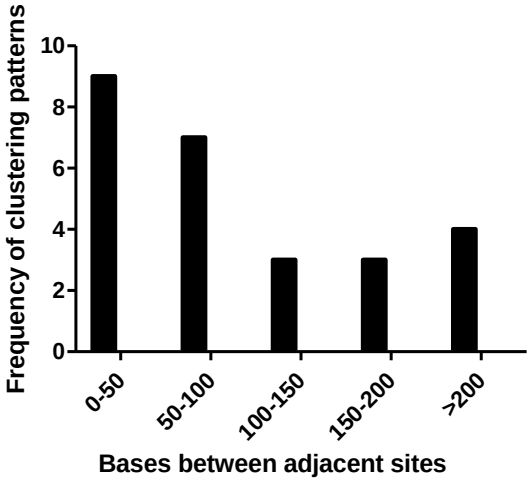
The *pap operon* is the most-well studied example of Dam's involvement in epigenetic gene regulation (2,3,6,7). The *pap operon* contains two GATC sites separated by 102 base pairs. These GATC sites both are within LRP (leucine-responsive regulatory protein) binding sites, where LRP is also involved in the regulatory process. When the operon is in the ON state the LRP is bound to the unmethylated GATC site distal to the *pap* gene; the GATC site proximal to the gene is fully methylated. When in the OFF state, LRP is bound to the unmethylated site proximal to the gene, blocking transcription. What remains unknown is mechanism for the transition between the ON and OFF state. In particular, we are interested in the transition from the unmethylated to the fully methylated GATC site. A direct transition between unmethylated and fully methylated to sites may rely on a intrasite methylation mechanism. Our work here attempts to address if this is possible in an *in vitro* system.

GATC sites known and postulated to be involved in gene regulation are highly clustered (Figure 5.2, Table 5.1) and have unique flanking sequences in comparison to the majority of genomic GATC sites (6). Some sites have a conserved A-tract 5' to the GATC site, which affects the methylation rate and the enzyme's intersite processive methylation (12, 13). An A-tract is referred to as a non-preferred flank, while most non-AT rich flanks are referred to as preferred. Intersite processivity refers to an enzyme's ability to modify two or more sites without dissociating. Other sites have AT rich flanks, which have modestly lowered methylation rates, and most have A-tracts near and around the GATC sites. It appears as if these transiently unmethylated GATC sites are in similar DNA contexts—in the small minority of regulatory sites in the *E. coli* genome—distinguishable from the majority of other GATC sites. Upon more robust classifications, these sites may form a set of "molecular rules (14)," providing a basis for identifying new epigenetically regulated operons, and giving insight into the function of the other unmethylated GATC sites in the *E.coli* genome. We want to explore how Dam behaves at these unique regions.

Table 5.1. A list of known heritably unmethylated GATC sites (underlined) and the distances to their closest adjacent GATC sites. Most (19 out of 26) of these sites have been implicated to be involved in gene regulation because they overlap with binding sites for known regulatory proteins. Examples where the spacing between GATC sites is greater than 200 base pairs are not included. [a] sequence from *E. coli* k-12 chromosome MG155. nc 000913.2; "number" represents list from table in (1).

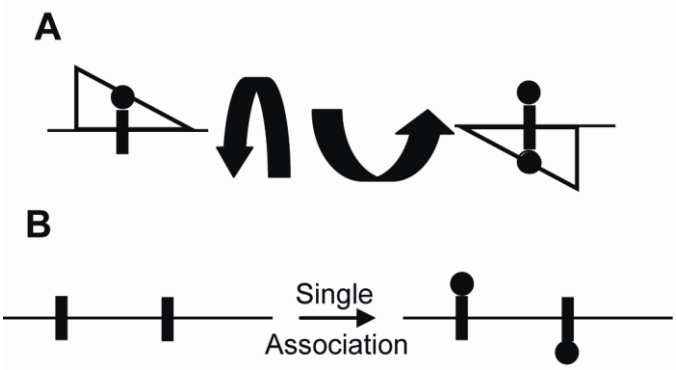
number; genome location[a]	distance between sites	sequence
2; 141280-141467	168	TAACATGTGT**GATC**GTCATCACAATTCGAGCTTTATTAACAGATTCCGCGAATGAATAGTTTTA CTGGTATACTGCGTGTCTTGCGCTTTGTTGCGGTGCCAAAACCTGCCCGTGCGAAGTGATTTGT TTTTAAATCATATGGTTAGAGATATGAAACATACTGTAGAAGTAAT**GATC**CCCGAAGCGG
3; 344331-344419	70	GATATTTTGTTGT**GATC**GGCGACACTTCGCTTAAAAAAGCACCAGTAGTGGTTTCGCAGCCATG CGGTGTATAAAAAAT**GATC**TCATGC
4; 621442-621662	197	TCGTTATC**GATC**TTATTTGGATATGTTAGCATGTGCAGCCTAAGAATAGGTATTTAAAATATTT GATGGCAAGGCATTGTAATGAATAAACAATCCTGGCTGCTTAACCTCAGCCTGTTGAAAACGCA CCCGGCGTTTCGCGCAGTATTCCTCGCTCGTTTCATCTCAATTGTGTCTCTGGGTTTGCTCGGC GTCGCGGTGCCGGTGCA**GATC**CAGATGAT
5; 765190-765268	55	GTGAAATT**GATC**ACATAATGGTATTGTTTTATCGGGCACACTGGCGCGACTATAAAAAC**GATC**A AGTGAG**GATC**ATGAT
6; 1099413-1099492	53	AATAAGTGT**GATC**TACGTCACTCATAACTGCAACGGATAATTTGTTGTTGCATAAAATGTGTGC TC**GATC**TCATTCATGG
7; 1168208-1168301	20	TGTATGT**GATC**CA**GATC**ACATCTATCATTTAGTTATC**GATC**GTTAAGTAATTGCTTGCGACGTC ATTCATCTGCATAAGGCCACTATTATGAAA
8; 1365999-1366177	169	TTCAATCA**GATC**TTTATAAATCAAAAAGATAAAAAATTGGCACGCAAATTGTATTAACAGTTCA GCAGGACAATCCTGAACGCAGAAATCAAGAGGACAACATTATGGGTATTTTTTCTCGCTTTGCC GACATCGTGAATGCCAACATCAACGCTCTGTTAGAGAAAGCGGAA**GATC**CA
9; 1653141-1653271	23	AAACTTCAGCCTTTAC**GATC**TTCATGTTCGATTCCTTGCATCGCTTGTCGTGATGCATGAAATC TACGCAACTGAGCTACTACCATACAAGTATAAA**GATC**GAAAAAAGCCGGAGT**GATC**ACAAAAAA AGG
10; 1859438-1859479	23	AAAACACAGATAAT**GATC**TGCGTTTTACAACTCA**GATC**ACAA
12; 2069331-2069399	22	AGAATAAAAC**GATC**AATATCTATTTTATC**GATC**GTTTATATC**GATC**GATAAGCTAATAATAACC TTTGT
13; 2229690-2229814	105	T**GATC**GCACCGTTCCTTTTCCCGATTATTCTGGCAGTAATGGGCTAAAATTTGCGATGCGTCGC GCATTTTTGATGTATGTTTCACGCGTTGCATAATTAATGAGATTCA**GATC**ACATATAAAGC
14; 2599023-2599032	2	CTCGT**GATC**AA**GATC**ACA
16; 3490434-3490522	22	GAC**GATC**ACTTTTTATTCCGT**GATC**AAAATCACCTCTTAAAATGCAATTTAGCAACCGATTGCA ATAAAACATTTAAACA**GATC**ACAAA
17; 3638634-3638751	96	GTTACCT**GATC**AGCGTAAACACCTTATCTGGCCTACGGTCTGCGTACGCAATCAAAATCCCCAG CCAATACAACATTTAACACCATCATATTTTCCATCATTAGTGT**GATC**ATCTGGT
18; 3740597-3740743	50	AC**GATC**TCGACAGATAATTTATAACCAATTGATTTTTATGTCTTTTGAAATTCATCAATCAGAT TGCCTTGTTAAAAAGT**GATC**GATATATTTGAAATCAAGTTTCGCATATTGAAATTTTAAGCCAA AAAAGC**GATC**AAAAAAACA
19; 3769939-3770003	39	AACATGCTGTA**GATC**ACATCAGGTGAACGCCGTAAGAAAATATCTTGTGATTCA**GATC**ACAAAG A
20; 3873223-3873339	98	CCC**GATC**GATGCATAACGCGGCAACTTTGTAGTACCAGCGTGATGACGTTCGCGTTTGCCGTGC GTGTAATGTAGTACAAACTTATATTGTTGTACTACAATTTA**GATC**ACAAAAAG
21; 4071646-4071805	133	TGCTTGTCTGTTTTT**GATC**GTATTTGTAATTTATCGTCAAAAAATTGACAGCCGTCACTTTTTA AACAATTGGTGAAATTAAAATGAACGCATCCCAAAATGTTTAAGGAATGACCATGATTCGTGTT GCTTGTGTAGGTATAACCGTGATG**GATC**GCAT
22; 4099532-4099618	38	TTTTTGT**GATC**AATTTCAAAATAAAAACAAT**GATC**CGATAAAAATAAAACAGCGTTTCAATTGA TGTGGTTTT**GATC**ACTTTTATTG
24; 4347090-4347259	138	AGATTAATCT**GATC**TACCCATTTGTGGGTAAAAATACACATAATGCGGGTGACATAATAGTTAA TTAACTTTTGTTAGCGTTTTGAAATTAAAAACACCGTTCACCTGAAGAGATATTAATTTTTAGC GATGATGGAGGGATAATTATATTT**GATC**TGGCACAAGTTTTA
25; 4537910-4538119	59	ACCTGTTATACCA**GATC**AAAAATCACGCAATCCATACAACAAAACCAGATTTGCAATTCGTGTC ACAAAATATGTC**GATC**TTTTTCTAAGAGGAAGATGCCATGTGAAGCCAGACGAACACTTGCGGT GGTCTTCAAAAACTAAA**GATC**TTAGTTTAACTATTTGT
26; 4537959-4538108	65	TTATACCA**GATC**AAAAATCACGCAATCCATACAACAAAACCAGATTTGCAATTCGTGTCACAAA ATATGTC**GATC**TTTTTCTAAGAGGAAGATGCCATGTGAAGCCAGACGAACACTTGCGGTGGTCT TCAAAAACTAAA**GATC**TTAGTT

Figure 5.2. Histogram of variable clustering of GATC sites (compiled from Table 5.1).



Initial studies on unmethylated DNA showed that Dam is able to methylate the adenines on both DNA strands within a single cognate site before dissociation, which was referred to as intrasite processivity (15). To accomplish this, the enzyme must switch strands and reorient itself, breaking and reforming its contacts with the DNA (Figure 5.3A). The restriction endonucleases BfiI (16) and BcnI (17), which cleave both strands of DNA within one cognate site, appear to rely on a similar reorientation of a single active site. The phenomenon of intrasite processivity is suggestive of hopping, where proteins interact with and move along DNA not only by 1-D sliding, but by using several dissociation-reassociation steps (18,19). While diverse techniques have been used to show that several proteins rely on mechanisms other than sliding (20,21), there are limited details about how hopping works.

Figure 5.3. Schematic of intrasite and intersite processivity: (A) Schematic of Dam's (◺) intrasite processivity where each adenine in the palindromic GATC site is methylated. Arrows represent how Dam must both switch strands and rotate 180 degrees. (B) Intersite processivity experiment where enzyme encounters a piece of DNA and can modify sites on both strands before dissociation. See text and references for diversity of enzymes and modifications. Sites to be modified are represented as (▌), and modifications to the sites are represented as (●).



Hopping has generated significant interest as a way to understand how proteins can efficiently find their recognition sites when an overwhelming excess of non-specific DNA is present (22). Hopping also has been used to explain how an enzyme can processively modify DNA when two or more sites have the opposite strand orientation, requiring the enzyme to reorient itself between modifying the first and subsequent sites (Figure 5.3B). Importantly, during this process the enzyme has a greater probability of rebinding the original DNA molecule than binding to another DNA molecule. Several enzymes display this activity, including T4 Dam (23), uracil DNA glycosylase (24), human alkyladenine DNA glycosylase (25), and BbvcI restriction endonucleases (26). Given the diversity of enzymes that can switch DNA strands and the general lack of mechanistic understanding of the underlying processes, we sought to explore the factors which regulate Dam's intrasite hopping mechanism.

Our original description of intrasite methylation by Dam relied on short, single site

synthetic double stranded DNA. In contrast, prior work with plasmids showed this activity is largely suppressed (1), suggesting that flanking DNA segments longer than those used in our original studies may regulate this activity. While the majority of GATC sites on the bacterial genome are predicted to be separated by ~260 base pairs, GATC sites known and postulated to be involved in gene regulation are generally separated by ~10 to 100 base pairs (Figure 5.2). The purpose of this study is to characterize how the sequence contexts of GATC sites—specifically their clustering—regulate Dam's intrasite processivity, with the goal of better understanding the mechanisms of Dam's varied roles in the cell. We also want to explore how proteins are able to processively modify their cognate sites by switching strands.

5.3 Results

5.3.1 Assay development

The original evidence for intrasite processive methylation relied on single turnover experiments and a tritiated AdoMet assay with a single site 21 base pair double stranded DNA substrate (15). For Dam, product release is rate limiting, and the observed rate constant from a single turnover reaction is the methyl transfer rate constant, k_{chem}. (27) Our experiments are defined as single turnover because enzyme is in excess of DNA molecules—but not available adenines—and the addition of more enzyme does not alter k_{chem} (data not shown). The tritium assay measures total methylation, which can be both single and double methylation (of a single site, which has two adenines); it cannot be used to directly monitor the formation of double methylation. Here, we sought to track double methylation events exclusively. To address this, we developed an assay using the restriction endonuclease DpnI, which cuts doubly methylated GATC sites significantly more efficiently

than hemimethylated sites. We optimized conditions so that no hemimethylated DNA was cut and >85% of doubly methylated DNA was cut (Materials and Methods, Figure 5.4). This was done to confirm that the reaction defined as intrasite processive involved no hemimethylated intermediates. To validate this assay, the DpnI cutting profile was compared to the tritium assay, giving the same rate constant (Figure 5.5A). This confirms that for this substrate (1B, see Table 5.2) both experimental methods, which are identical except for the read-out, measure the rate of double methylation, the action defined as intrasite processivity.

Figure 5.4. Validation of the DpnI cutting assay. (A) Substrate H is methylated by M.Taq to create substrate Hhm (hemimethylated GATC site on DNA). Lane 1 is substrate Hhm (hemi methylated 60mer) subjected to DpnII digestion. DpnII cuts unmethylated GATC sites. Lane 2 is substrate H subjected to DpnII digestion (digestion fragments are both ~30 base pairs and are shown overlayed on the gel). Protection from cutting shows that substrate Hhm is completely hemi methylated. (B) Substrate Hhm and substrate Hdm (completely methylated GATC site on DNA) are subjected to DpnI digestion. Lane 1 shows that substrate Hhm is uncut while substrate Hdm is significantly cut.

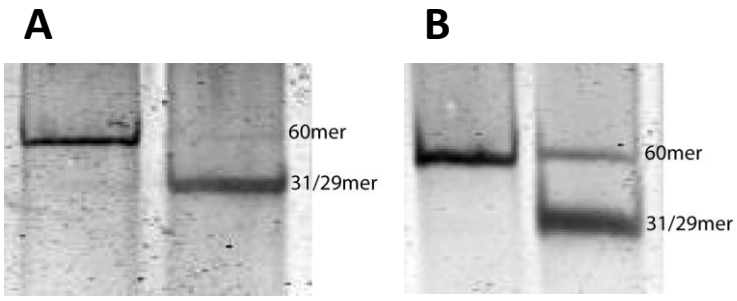
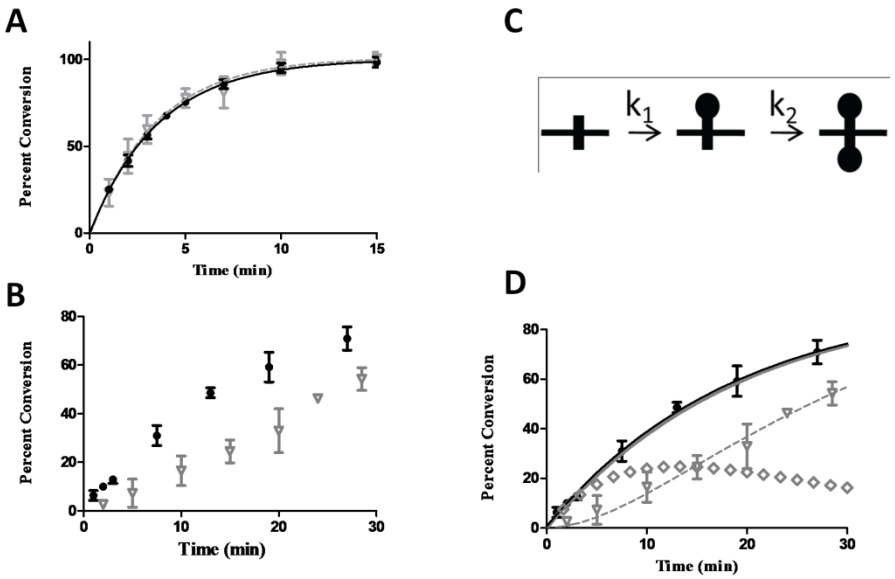


Figure 5.5. Validation of DpnI assay to detect intrasite processivity. Comparison of tritium and DpnI assay, 200nM DNA, 210nM Dam, 30μM AdoMet, and 15 degrees. 100 Percent conversion for DpnI trace refers to complete methylation of substrate, and subsequent DpnI digestion based on assay described in Figure 5.4. Tritium data is represented by black dots (●), and DpnI data for the 115 bp fragment is represented by grey triangles (▽). Error bars represent between 2-5 replicates. (A) Substrate 1B: single turnover fit for tritium is a solid black line; and single turnover fit for the DpnI data is a dashed grey line. (B) Tritium data (●) and DpnI data (▽) for substrate 1C. (C) Reaction scheme for total methylation of an unmethylated site. k_1 is the methylation rate constant from unmethylated to hemimethylated; k_2 is methylation rate constant from hemimethylated to doubly methylated. (D) Rate constants k_1=0.10 min^{-1} and k_2=0.053 min^{-1} are used in equation 3 (dashed grey line) and equation 4 (solid grey line) to fit with the data. Also included is the single turnover fit from the tritium data (solid black line), and the profile for hemimethylated DNA (grey diamonds).



Dam does not always display intrasite processivity, as demonstrated by a delay in the DpnI cutting profile in comparison to the tritium data (Figure 5.5B, substrate 1C, Table 5.2). The discrepancy in the observed activity of the two methods is attributed to the presence of a hemimethylated intermediate, which is enumerated in the tritium assay, but not by DpnI. Using the tritium data, the DpnI data, and kinetic modeling, we confirmed that this delay

represents an almost completely non-intrasite processive mechanism, which will be referred to as sequential. For the sequential reaction, Dam methylates one strand, then completely dissociates from the DNA before returning to methylate the second strand. Dam's ability to fully methylate a GATC site can be simplified into the reaction scheme in Figure 5.5C, and the individual components can be tracked by equations 2-4 (k_1 and k_2 are as described in Figure 5.5; A_0 is the plateau level, which is 100%) regardless of its methylation mechanism. These equations are used to relate the tritium data to the DpnI data, both of which are read-outs for the same reaction. While k_1 and k_2 represent observed rates of methylation, imbedded in each term are several other events, such as DNA binding, translocation, and methylation. Several groups have used the strategy employed here of directly monitoring activity, not including the other microscopic rate constants, to model processive, non-processive, and partially processive events (24, 26). Processivity is defined simply as the relative enhancement in k_2 over k_1. Knowing the values of k_1 and k_2 allows one to profile individual species of the reaction separately, showing the amount of single methylation, double methylation, and total methylation (which is a combination of single and double methylation) at any time point. The double methylation product profile is equivalent to the DpnI data. Total methylation, reflecting the sum of hemimethylation (B, Equation 2) and double methylation (C, Equation 3), is equivalent to the tritium data (Equation 4). Therefore, k_1 and k_2 can be modeled using a least-squares fit such that the tritium data matches the trace from Equation 4 and the DpnI data matches the trace for Equation 3 (Figure 5.5D). In summary, the legitimacy of the DpnI assay is confirmed by deriving rate constants from the DpnI data and matching the trace with the tritium data.

$$\text{Hemimethylation (B)} = (A_0 k_1/(k_2-k_1))(e^{-k1t} - e^{-k2t}) \quad \textbf{(2)}$$

Double methylation (C) =

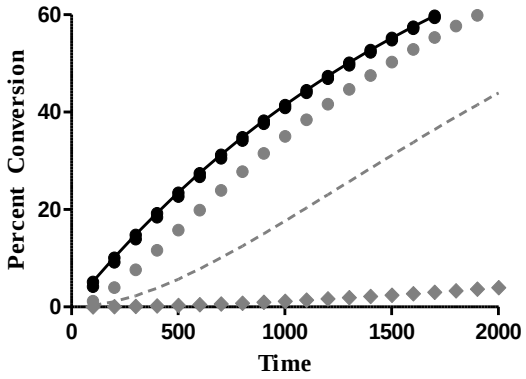
$$A_0[1+(1/(k_1-k_2))(k_2e^{-k1t} - k_1e^{-k2t})] \quad \textbf{(3)}$$

Total methylation = B + 2C **(4)**

For an intrasite processive event, k_2 is much faster than k_1; the initial methylation is followed by a rapid methylation of the opposite strand with no detectable hemi-methylated products. The enhancement in k_2 comes from the enzyme maintaining contact with the DNA during both methylations, foregoing dissociation and rebinding steps. Alternatively, for a sequential reaction, each methylation event involves free enzyme poised to bind DNA, and the first methylation event would not affect the second methylation. Since the rate of the methylation of hemi- to fully-methylated is close to that of non-methylated to fully methylated, k_1 would be predicted to be similar to k_2 for a sequential process (15). Since enzymes can display a gradient of processivity, several ratios of k_2:k_1 were modeled to predict the DpnI traces of partially processive scenarios (Figure 5.6). Notably, when k_2 is only 10 fold larger than k_1, the DpnI trace matches the tritium trace. This suggests that there is a narrow window of possible rate constants between sequential and intrasite methylation. A sequential process requires the enzyme to undergo product dissociation, the rate limiting step, and release of the cofactor product before the second methylation step can occur. These processes could make k_2 slower than the observed rate of k_1.Given this reasoning, it is impossible to define what ratio of k_1:k_2 would constitute a truly sequential process. Interestingly, the tritium data fits to a single exponential (one observable rate constant) for both sequential and intrasite methylation. This observation can be rationalized: If the site is methylated entirely via an intrasite mechanism, the second rate constant is too fast to be detectable; when sequential, the two rate constants approach each other, again resulting in

what appears to be a single exponential.

Figure 5.6. Traces for different ratios of k_1:k_2 used to model the data from Figure 5.3D.The black line is the tritium data; the black dots are k_2=1000k_1 and k_2=100k_1(both traces are identical); the grey dots are k_2=10k_1; the grey diamonds are k_2=.1k_1; the grey dashed line is sequential methylation, k_1~k_2. Notably, only a small ratio of rate constants separates a sequential from an intrasite processive event. To address what a truly sequential process might be, the scenario k_2=.1k_1 was modeled, showing a large decrease in activity in comparison to the tritium data and the k_1~k_2 scenario.

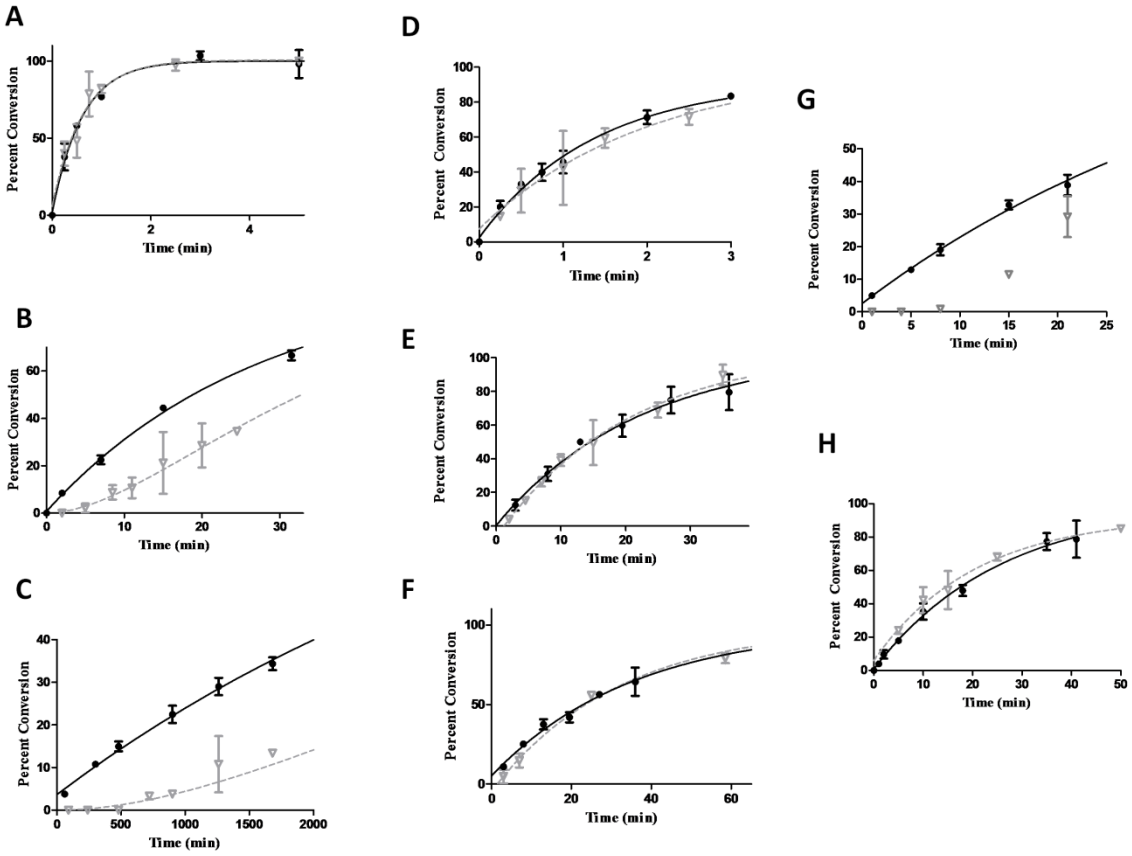


5.3.2 WT single turnover results

Intrasite processivity was confirmed on a short synthetic piece of DNA (substrate 1A) with non-preferred flanks (Figure 5.7A). Adding additional flanks of ~100 base pairs (PCR derived substrates, see Methods and Materials) to each side of the GATC site still showed intrasite processivity (Table 5.2, Figure 5.5A). However, when extending the flanks on one side of the site, the DpnI cutting data matches the trace for a sequential mechanism. Again, k_1 and k_2 are modeled such that the tritium data correlates with the Dpn1 data (Figure 5.3). This shows that intrasite processivity is compromised by flanking DNA (Figure 5.7). As seen in Table 5.2, greater than ~400 base pairs of non-specific flanking DNA on one side of the GATC site causes the reaction to be sequential for single site substrates. Additionally, the methylation rate constants decrease significantly upon the addition of larger flanks (see

discussion).

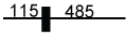
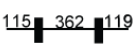
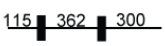
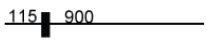
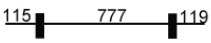
Figure 5.7. Intrasite processivity is modulated by lengths of flanking DNA. The tritium data (●) and DpnI data (▽) for substrates from Table 5.2. k_1 and k_2 (as described in the text and in Figure 5.5, min^{-1}) are supplied for the single site sequential substrates (A) substrate 1A, (B) substrate 1D; k_1= 0,072, k_2= 0.036 (C) substrate 1E; k_1= 0,025, k_2= 0.014 (D) substrate 2B, (E) substrate 2C, (F) substrate 2D, and (G) substrate 2E, (H) substrate 1LL.While the kinetic scheme is too complicated to predict what sequential methylation of each site would be for substrate 2E, the characteristic delay in the DpnI trace in comparison to the tritium trace is convincing enough to assume the hemimethylated intermediate is present. For (D-G) the DpnI data represents the accumulation of the 115mer fragment, which was nearly identical to the 119mer fragment (not shown).



Double site substrates were used to explore the relationship between intrasite processivity and the clustering of GATC sites. The reaction conditions were the same as used for single site substrates (Enzyme:DNA, 1.05:1); note, here the number of GATC sites is ~twice the amount of enzyme. Interestingly, all four methylation events were embedded in

a single rate constant for methylation by following tritium incorporation (Figure 5.7D-F). The DpnI data, which is simpler to interpret than the tritium because it traces each site separately, showed that each site underwent intrasite processivity. This was shown for several substrates (Table 5.2). However, intrasite processivity is compromised when 777 base pairs are added between the two sites (Table 5.2, Figure 5.7G).

Table 5.2. Intrasite processivity is modulated by lengths of flanking DNA. The numbers under the "cartoon" column refer to amount of base pairs surrounding the GATC sites. k_{chem} refers to the observable single exponential rate constant (min^{-1}) derived from the tritium assay. "Yes" and "No" under the "Intrasite" column refer to intrasite processive and sequential respectively. See Figure 5.7 for tritium and DpnI assays of these substrates.

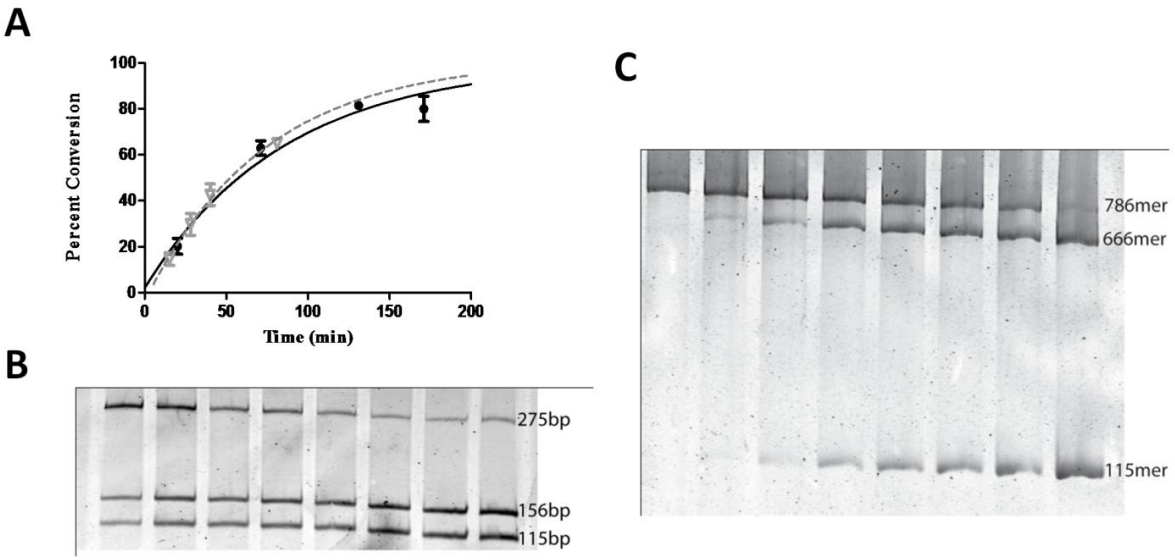
name	cartoon	k_{obs} (min^{-1})	intrasite?	name	cartoon	k_{obs} (min^{-1})	intrasite?
1A		1.8 ± 0.1	Yes	1LL		0.039± 0.01	Yes
1B		0.29 ± 0.06	Yes	2B		0.063± 0.06	Yes
1C		0.052 ±0.005	No	2C		0.046 ± 0.005	Yes
1D		0.036 ±0.002	No	2D		0.028 ± 0.003	Yes
1E		0.014 ±0.001	No	2E		0.023 ± 0.002	No

The DpnI assay allows for a direct comparison of methylation profiles for substrates with and without an additional GATC site. Based on the information in Table 5.2, only if more than ~400 base pairs surround the site is intrasite processivity lost. Introduction of a second GATC site, however, resurrects intrasite processivity at both GATC sites. The third and fourth rows of Table 5.2 show identical substrates, except the ones on the right have an additional site added and show intrasite processive methylation. Additionally, a single site substrate, 1LL, with flanks of 300 and 350 base pairs surrounding the GATC site is intrasite

processive (Figure 5.7, Table 5.2).This shows that decreases in methylation rate constants alone do not lead to a sequential mechanism because substrate 1C has a similar rate constant but does not display intra-site processivity.

Taken together, the single and double site data show that intrasite processivity is sensitive to the amount of DNA surrounding GATC sites. Also, the results in Table 5.2 can address the potential concern that intrasite processivity manifests itself by Dam switching strands at the ends of the DNA. Substrate 1B shows intrasite processivity, where the shortest distance from the GATC site to the end of the DNA is 115 base pairs; substrates 1C-E have the same feature, but are not intrasite processive. Therefore, if the "end-effect" contributed to intrasite processivity, substrates 1C-E would have some intrasite character, which is not shown. Furthermore, the low concentrations of DNA do not support the functional oligomerization of Dam (15).To further explore how flanking non-specific DNA affects the observed methylation rates and intrasite processivity, a competition experiment was done using substrate 1A and the same amount (in molecules) of a 500 base pair piece of non-specific DNA (Figure 5.8). Substrate 1A retained intrasite processivity, confirming that the non-specific DNA needs to be flanking the site to be inhibitory. Additionally, the observed rate constant decreases significantly (k_{chem} is 0.011+/- .001 min^{-1}) in comparison to when no competitor DNA is present. This outcome suggests that Dam spends an appreciable amount of time on the non-specific DNA. If Dam has no affinity for non-specific DNA, then the addition of competitor DNA would be predicted to have no effect on the rate of methylation (discussion).

Figure 5.8 Competition experiment with non-specific DNA. (A)The tritium data (●) and DpnI data (▽) for substrate 1B with a 500 base pair piece of chase DNA included. Single exponential fits for tritium and Dpn1 is a black line and a dashed grey line respectively. k_{chem} for reaction is 0.011+/- .001 min^{-1} (B) Representative gel for the reaction time course for substrate 1B (Table 5.2). The DpnI data is derived from the accumulation of the 115mer base pair fragment. The final lane represents the plateau level from the complete methylation of the substrate. (C) Representative gel for substrate 1C (Table 5.2).

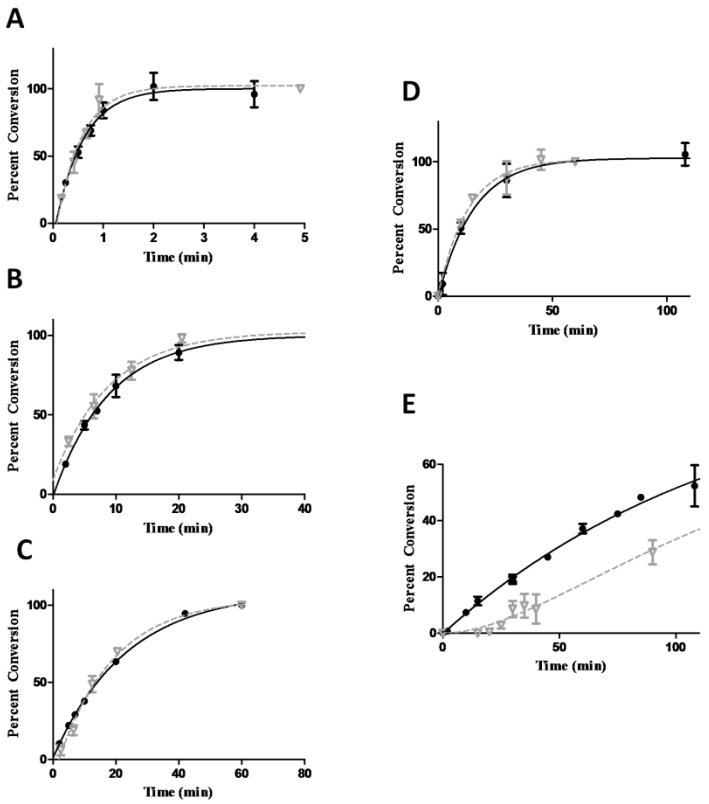


5.3.3 Mutant single turnover results

Our previously characterized Dam mutants are disrupted at the interface between Dam and DNA outside of the GATC site; these mutants uniformly show decreases in methylation rate constants, and in intersite processivity (29). The latter result implies that disruption of this interface challenged Dam's ability to move on DNA. Dam's movement on DNA, including its transition from specific to non-specific sites and its ability to switch strands, is a fundamental aspect of intrasite processivity. Therefore, we explored each mutant's potential for intrasite processivity (Figure 5.9). All but one mutant shows intrasite processivity on a short synthetic oligonucleotide with preferred flanks; N126A, the mutant

with the largest reduction in methylation, was shown to undergo a sequential reaction scheme. This finding shows that a single DNA-contacting residue can disrupt intrasite processivity, and provides insight for WT Dam's intrasite processive mechanism (discussion).

Figure 5.9. Intrasite Processivity of Dam mutants. The tritium data (●) and DpnI data (▽) for Dam mutants on a 60 base pair substrate. (A) K139A, (B) N132A, (C) R116A, (D) R95A, (E) N126A. Only (E) is sequential.



5.4 Discussion

The mechanism whereby epigenetically controlled operons switch methylation states to facilitate bacterial gene regulation remains elusive (Figure 5.1). While several rounds of replication are necessary for the transition from doubly to unmethylated sites (no *E.coli*

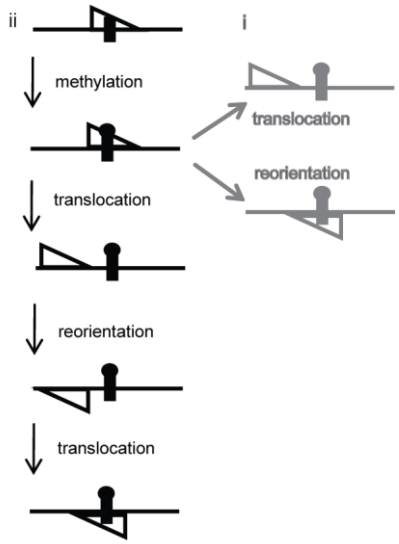
adenine demethylase has been identified), the opposite transition may have no restrictions. Intrasite processivity could be a way for unmethylated sites to be fully methylated in one round of replication, since there are very low levels of Dam in the cell (30), allowing the cell to respond more efficiently to external stimuli. The unmethylated GATC sites known and speculated to be involved in epigenetic gene regulation all have unique contextual properties in comparison to the overwhelming majority of GATC sites which are involved in the mismatch repair system. In this study we show that intrasite processivity is dependent on the context of the GATC sites, providing a basis for the modulation of Dam activity at particular genomic locations. The mechanistic effects of the clustering of methylation sites seen here may be related to a broader biological paradigm. For example, methylation sites in promoters modified by the human *de novo* DNA methyltransferase DNMT3A are overrepresented by 5 fold in CpG islands (31). This clustering, along with flanking sequence preferences, modulates DNMT3A's processivity (32).

Our data shows that when presented with DNA which is hundreds of basepairs in length, Dam carries out double-methylation of an unmethylated site only in the presence of another site, less than 400 bp away. This, along with the activity of other highly regulated DNA binding proteins provides insight into potential mechanisms of epigenetic gene regulation. Under our experimental conditions, intrasite processivity is facilitated by other sites as far away as ~400 base pairs, which is much more than necessary to accommodate the separation of ~10-100 base pairs observed at genomic regions undergoing epigenetic regulation (see introduction). However, under physiological conditions (salt, molecular crowding, ect.) Dam's ability to stay associated to DNA between sites may be hindered.

To doubly methylate DNA, Dam must both switch strands and flip its orientation (Figure

5.3).Two potential mechanisms seem plausible: (i) the enzyme dissociates and reassociates directly at the GATC site as it switches strands, and (ii) the enzyme dissociates to the DNA flanking the GATC site as it switches strands (Figure 5.10). Mechanism (i) involves a partitioning between direct reorientation and movement onto adjacent nonspecific DNA. Importantly for mechanism (i) this movement to nonspecific DNA is not an intermediate for intrasite processivity. In contrast, mechanism (ii) proposes an indirect reorientation process with an obligate binding to adjacent nonspecific DNA to achieve intrasite processivity. Three independent lines of evidence support mechanism (ii). First, the observed methylation rate constant decreases with increasing DNA, as does the extent of the reaction. This is inconsistent with mechanism (i), even if the enzyme moving to a flanking DNA segment is in competition with the direct dissociation and reassociation mechanism. In other words, mechanism (i) predicts a decrease in the asymptote but not in the rate constant for intrasite methylation. Second, intrasite processivity is salt dependent, where greater than ~150 mM NaCl inhibits the enzyme from methylating both strands at once (15). A salt dependency in processivity is consistent with hopping (25) due to the effect on binding and dissociation. Therefore, Dam most likely uses hopping to accomplish intrasite processivity. While the precise nature of hopping has not been defined for any protein, it involves some dissociation of the enzyme from the DNA which results largely in the reassociation to the same DNA, but at sites removed from the originally bound site. Hopping has been estimated to account for protein "jumps" as far as 100 base pairs of DNA (33). The re-association for intrasite processivity, therefore, will most likely be at a non-specific site.

Figure 5.10. Potential models of intrasite processivity and its regulation. The direct mechanism (i, from the text) is depicted in grey. The indirect mechanism (ii, from the text), where intrasite processivity proceeds by an intermediate with flanking DNA, is shown in black. Notably, the translocation step in (i) represents a loss of intrasite processivity. For (i) the observed rate of the reaction and the occurrence of intrasite processivity would not be predicted to change with increases of flanking DNA.

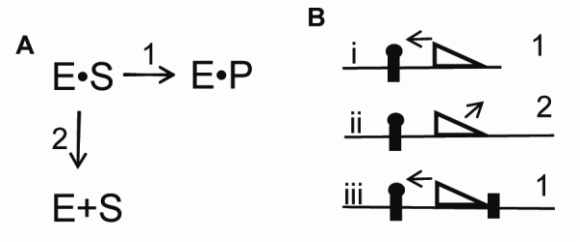


Third, our observation that intrasite processivity is lost when only one side of the GATC site has flanking DNA beyond 400bp suggests that the enzyme samples both sides of the GATC site in its path towards intrasite methylation, further arguing that Dam binds to non-specific flanking DNA when carrying out intrasite processive catalysis. If this were not the case, one would expect that intrasite processivity would be decreased only if both sides have flanks longer than ~400 bp, which we do not observe. In other words, Dam must traverse the single GATC site at least once when first moving in the direction of the shorter segment (115bp) to display a complete loss of intrasite processivity. Furthermore, our interpretation of the mechanism of intrasite processivity, namely the role of flanking DNA, was similarly proposed by Siksnys *et al.* for the intrasite processive BcnI restriction endonuclease (see introduction) (17).

The basis for this conclusion relied on an experiment with a catalytically inactive mutant BcnI enzyme chase (17). Following pre-incubation with the active enzyme and the DNA, the inactive chase enzyme is added in excess to the reaction simultaneously with the Mg^{2+} cofactor, which is required for catalysis. The first cutting event proceeds as normally, but the second cutting event does not occur. This data suggests that the following the first cutting event, the enzyme leaves the site such that the inactive mutant can replace it. A following experiment suggests that when leaving the site the enzyme associates with the flanking DNA, similar to our interpretation here. In this following experiment excess chase DNA is added instead of the inactive mutant. Here, the second cutting event proceeds more effectively than in the case with the inactive mutant chase. If the enzyme hops a significant distance away from the DNA, then the excess chase will capture it, which is not the case. Instead, the authors argue that the enzyme likely makes a short hop to the adjacent DNA, and then able to return to the site to cut the other strand. These results lend precedent to our model in Figure 5.10

The ability of an adjacent Dam recognition site to reinstate intrasite processivity on DNA with large flanking sequences is at first difficult to understand (Figure 5.11). However, others have demonstrated that the lifetime of a protein on a segment of DNA is enhanced when multiple recognition sites are placed within certain distances (34). Longer retention of the protein on the DNA results from the protein translocating to the tight binding sites. We hypothesize that by providing Dam these adjacent recognition sites, the enzyme has greater opportunity to return to the original site and carry out intrasite processive catalysis.

Figure 5.11. How Flanking DNA regulates intrasite processivity. (A) "E" is Dam, "S" is hemimethylated DNA, and "P" is fully methylated DNA. Shown is a schematic depicting the possible outcomes following the initial methylation of a GATC site. Either the enzyme will undergo intrasite processivity (scheme 1), or the enzyme will leave the hemimethylated substrate (scheme 2). (B) The Type of substrate dictates which mechanism occurs from A (1 or 2). In (i) the enzyme stays associated with the DNA long enough to remethylate it. However, in (ii) the enzyme leaves the DNA because it spends too much time on the non-specific DNA away from its GATC site, forcing scheme 2. In (iii) the second GATC site allows Dam to spend longer on the DNA, pushing the reaction towards (1).



Dam alanine mutations made to phosphate contacting residues adjacent to the recognition site showed the interesting result of decreased methylation and inter-site processivity, but K_d dissociation constants that remained similar to WT (29). Here we show that N126A Dam, the mutant with the most severely decreased methylation and inter-site processivity, is the only mutant incapable of carrying out intra-site processivity (Figure 5.9). Our previous explanation for why the processivity of N126A and the other Dam mutants are decreased invoked the decreased methylation rate at the second site (29). This is most likely occurring with N126A during intra-site processivity as well. Thus, the slower methylation combined with the unchanged affinity should result in decreased intra-site processive methylation. What is intriguing is that all of the originally studied Dam mutants showed some decrease in inter-site processivity, whereas the most severely impacted mutant (N126A) shows changes in intra-site processivity. This may derive from the use of preferred and non-preferred flanking sequences for the prior, inter-site study (29), whereas here the single site has a preferred flanking sequence. The modeling results (Figure 5.6) suggest that

our assay for intra-site processivity is extremely responsive to small changes in the relative values of the two methylation events, whereas the assay for inter-site processivity is less responsive.

5.5 Materials and Methods

DNA substrates: All restriction endonucleases were obtained from NEB. All synthetic DNA substrates and primers were obtained from IDT and Midland and were re-suspended in TE buffer (10mM Tris, pH 7.5, 1mM EDTA). They were annealed with their reverse complements in a 1:1 mixture for 5 minutes at 95 degrees, and allowed to cool to room temperature (~5 hours). Annealing was verified by PAGE. Substrate 1A (Table 5.2) is 5'-GTTCGTCATGCATGCAATGGAAAAGATCAGGTACCTGAATCACGAACGTTAGGCATTCGC. The substrate used in the mutant anaylsis (Figure 5.9) is: 5'-ATCGTGGACTTCTACTTGGATGGAGGATC GGATGACACGTATTCCAGGAATTCACGTTAC-3'. The production of several PCR amplicons used the following strategy: A synthetic oligonucleotide with two GATC sites, and two restriction sites between the GATC sites was cloned into plasmid pBR322 (NEB). 362 and 777 base pair spacers were generated by PCR and cloned into the plasmid, generating different distances between the GATC sites. These plasmids were PCR amplified with different primers to adjust the spacings from the GATC sites to the ends of the DNA. The same strategy applied to an engineered vector with a single GATC site.

The following substrates were cloned into the plasmid pBR322 at the EcoRI and HindIII sites. Double site: 5'-AATTCGGTGATCTTTTCGA*C CCGGG*AGCTGGTAGTATGC*CCATGG*TTCGATCTTTT GCCA-3', and single site, 5'AATTCG GTGATCTTTTCGA*CCCGGG*AGCTGGTAGTATGC*CCATGG*

TTCGGTCTTTTGCCA -3', making new plasmids called pBRMut0double and pBRMut0single. The cloned, synthetic insert had additional cloning sites within it: XmaI and NcoI (italicized). These sites were used to insert PCR purified spacers between the two GATC site(s) (underlined). Upon PCR amplification, the spacers were digested with XmaI and NcoI to generate overhangs. The spacers were generated by PCR from the plasmid pBR322 with restriction sites using the same forward primer: 5'- ATT *CCCGGG* GGCTACCCTGTGGAACACCT - 3',with different reverse primers for each sized spacer: (2C, 2D from Table 5.2) 5'-TAAT*CCATGG* GCAGCTGCGGTAAAGCTCAT -3', (2E)5'-TAAT *CCATGG*CATGTTCTTTCCTGCGT TATCCCC-3'.Plasmid pBRMut0 was digested and the spacers were inserted making plasmids pBRMut2 and pBRMut3. Amplicons with 115/119 base pairs flanking GATC sites were amplified from plasmid pBRMut0,2,3 using primers: (forward) 5'- GGGTTCCGCGCACATTTCCC-3' and (reverse) 5'-CCAGGGTGACGGTG CCGAGG-3'. Amplicons with 300 base pairs flanking GATC sites were amplified from plasmid pBRMut0,2,3 using primers: (forward) 5'- GCATCTTT TACTTTCACC AGCG-3', and (reverse) 5'-GGCT CCAAGTAGCG AAGCGAGC-3'. PCR amplicons were purified using the Agilent PCR clean-up kit and ethanol precipitated to achieve desired concentrations.

Single Turnover reactions: All single turnover reactions were done in MRB (100mM Tris, pH8.0, 1mM EDTA, 1mM DTT, 0.2mg/mL BSA) with 400nM DNA, 420nM Dam, 0.2mg/mL BSA, and 30 µM AdoMet (6Ci/mmol mixture of unlabeled and [^{3}H]methyl labeled, Perkin Elmer). Reactions were initiated with DNA with a total volume ranging from 60-80 µl and were done at 15 degrees. 8µl reaction fractions were quenched with 8µl 1% SDS. 14.5 µl quenched fractions were spotted on DE81 filter paper. The paper was washed

three times with a 50mM KH_2PO_4 buffer, once in 80% ethanol, once in 95% ethanol, and was dried in diethyl-ether; all washing steps were for five minutes. Papers were dried and submerged in BiosafeII scintillation fluid. Tritium levels were quantified using a Beckman Coultier LS6500 scintillation counter and converted to nM DNA product. Plateau levels of 100% were defined by the complete methylation available adenines in the reaction mixture. All single turnover tritium reactions were fit to a single exponential (Equation 5, A_0 is the plateau level, which is 100%).

$$\text{Percent Conversion} = A_0(1-e^{-kt}) \quad \textbf{(5)}$$

DpnI assay: 2.5µl of the single turnover (here, 30 µM of unlabeled AdoMet) assay was heat inactivated in 14.8 µl of 75 degree water for 20 minutes After slow cooling, 2µl of NEB buffer 4 was added and the mixture was incubated at 37 degrees for ~20 minutes. 0.7µl of DpnI was added NEB (14 units) and the solution was rapidly mixed. The cutting reaction proceeded for ten minutes at 37 degrees until it was heat inactivated in an 80 degree water bath for 20 minutes and slow cooled to room temperature for subsequent gel analysis. The reaction products were analyzed using PAGE (20%-5% 29:1 acrylamide:bisacrylamide, depending on substrate length) for 2 hours at 300 volts in TBE. Gels were stained with SybrAu and scanned on a Typhoon Phosphoimager (GE). Nucleic acids were quantified using the software provided with the Typhoon. The density of the different nucleic acid bands after several hours of reaction incubation (complete methylation), and subsequent digestion with the DpnI restriction endonuclease are defined as having the reaction being 100% complete (Figure 5.4) See Figure 5.8 for sample gels.

Competition experiment: The competition experiment consisted of a single turnover reaction with substrate 1A, to which an equimolar amount of a 500 base pair non-specific

(no GATC sites) piece of DNA was added. The reaction was initiated with a mixture of substrate 1A and the nonspecific DNA. The nonspecific DNA was generated by PCR using plasmid pBR322 as a template and the forward primer 5'- ATTCCCGGG GGCTACCCTGTGGAACACCT-3' and the reverse primer 5'- TAATCCATGGCCCGGCATC CGCTTACAGAC

Enzyme expression and purification: Dam was expressed and purified as previously described (27). In summary, Dam was over-expressed in XL2 Blue (Stratagene) *E.coli* cells grown at 37 degrees in LB media with 25ug/mL kanamycin and 12.5µg/mL tetracycline. After reaching an OD of 0.4-0.6, the cells were induced with 1mM IPTG and 0.05% L-arabinose and grown for 2 hours. The pellets were re-suspended in 40mL P11 buffer (50 mM potassium phosphate, pH 7.4, 10mM β–mercaptoethanol, 1mM EDTA, 1mM PMSF, 0.2mM NaCl, 10% glycerol) and lysed by sonication: 70% amplitude, 2 seconds on, 15 seconds off, total time 1 minute. Lysate was centrifuged for 60 minutes at 15,000 rpm at 4 degrees. Supernatant was loaded onto a 60 mL phosphocellulose (Whatman) column. The protein was eluted with a salt gradient from 0.2 and 0.8 M NaCl, and fractions with DAM were pooled and dialyzed in BS buffer (20mM potassium phosphate buffer, pH 7.0, 10mM β–mercaptoethanol, 1mM EDTA, 1mM PMSF, 10% glycerol). Upon overnight dialysis, the protein was loaded onto a 20-mL Blue Sepharose 6 FastFlow (GE Healthcare) column, and eluted with a salt gradient between 0-1.5 M NaCl. Fractions were pooled and flash frozen at -80 degrees. The protein concentration was determined using the extinction coefficient of $1.16 mg^{-1} cm^{-1}$ at 280 nm.

DpnI digestion control experiment: Single stranded DNA and its reverse compliment were ordered and annealed (5'-

ATCGTGGACTTCTACTTGGATGGAGAAAA**GA<u>TC</u>**<u>GA</u>CACGTATTCCAGGAATTCA CGTTAC-3') generating substrate H. The DNA was then subjected to M.Taq methylation. The 50 uL reaction contained 210 nM DNA, 100 μg/mL BSA, 30uM SAM, 5uL NEB Buffer 4, and 30 units of M.Taq (NEB); it was incubated for 3 hours at 65 degrees. The reaction was then subjected to a phenol chloroform extraction and then an ethanol precipitation, making substrtate Hhm. M.Taq methylates adenines in 5'-TCGA-3' sequences (underlined), creating a hemimetylated GATC site (bold). To show that the substrate was completely hemimethylated, the DNA was subjected to DpnII (NEB) digestion. The hemimethylated DNA was then subjected to Dam methylation under single turnover reaction conditions (see primary manuscript) for several hours to insure complete methylation, making substrate Hdm. The Hhm control lane in S Figure 5.4B was incubated with the same single turnover conditions as Hdm, but no AdoMet was added. 2.5μ L of each reaction was added into 14.8 μL of water. 2 μL of NEB Buffer 4 was then added. Then 0.7 μL (14 units) of Dpn1(NEB) was added and the reaction was incubated at 37 degrees for 10 minutes. The reactions were heat killed and slow cooled to room temperature before PAGE.

5.6 References

1. Herman, G. E. & Modrich, P. (1982). *Escherichia coli* dam methylase. Physical and catalytic properties of the homogeneous enzyme. *J. Biol. Chem.* **257**,2605-2612

2. Wion, D. & Casadesus, J. (2006) N^6-methyl-adenine: an epigenetic signal for DNA–protein interactions *Nat. Rev. Microbiol.* **4**, 183-192

3. Marinus, M. G. & Casadesus, J. (2009) Roles of DNA adenine methylation in host–pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol. Rev*. **33**, 488-503

4. Touzain, F., Petit, M. A., Schbath, S. & El Karoui, M. (2011) DNA motifs that sculpt the bacterial chromosome. *Nat. Rev. Microbiol*. **9**, 15-26

5. Broadbent, S. E., Davies, M. R. & van der Woude, M. W. (2010) Phase variation controls expression of *Salmonella lipopolysaccharide* modification genes by a DNA methylation-dependent mechanism. *Mol. Microbiol.* **77**, 337-353

6. Casadesus, J. & Low, D. (2006) Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* **70**, 830-856

7. Lobner-Olesen, A., Skovgaard, O. & Marinus, M. G. (2005) Dam methylation: coordinating cellular processes. *Curr. Opin. Microbiol.* **8**, 154-160

8. Ringquist, S. & Smith, C. L. (1992) The *Escherichia coli* chromosome contains specific, unmethylated dam and dcm sites. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4539-4543

9. Tavazoie, S. & Church, G. M. (1998) Quantitative whole-genome analysis of DNA-protein interactions by in vivo methylase protection in *E. coli*. *Nat. Biotechnol*. **16**, 566-571

10. Hernday, A. D., Braaten, B. A. & Low, D. A. (2003) The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol. Cell* **12**, 947-957

11 Kaminska, R. & van der Woude, M. W. (2010) Establishing and maintaining sequestration of Dam target sites for phase variation of agn43 in *Escherichia coli*. *J. Bacteriol.* **192**,1937-1945

12. Coffin, S. R. & Reich, N. O. (2008) Modulation of *Escherichia coli* DNA methyltransferase activity by biologically derived GATC-flanking sequences. *J. Biol. Chem.* **283**, 20106-20116

13. Peterson, S. N. & Reich, N. O. (2006) GATC flanking sequences regulate dam activity: evidence for how dam specificity may influence papi expression *J. Mol. Biol.* **355**, 459-472

14. van der Woude, M. W. (2011) Phase variation: how to create and coordinate population diversity. *Curr. Opin. Microbiol*. **14**, 205-211

15. Coffin, S. R. & Reich, N. O. (2009) *Escherichia coli* DNA adenine methyltransferase: intrasite processivity and substrate-induced dimerization and activation. *Biochemistry* **48**, 7399-7410

16. Sasnauskas, et al. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′–5′ and 5′–3′ strands by rotating a single active site. *Nucleic Acids Res.* **38**, 2399-2410

17. Sasnauskas, G., Kostiuk, G., Tamulaitis, G. & Siksynys, V. (2011) Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.* **39**, 8844-8856

18. Halford, S. E. & Marko, J. F. (2004*)* How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040-3052

19. Stanford, N. P., Szczelkun, M. D., Marko, J. F. & Halford, S. E. (2000) One-and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* **19**, 6546-6557

20. Tafvizi, A., Mirny, L. A. & van Oijen, A. M. (2011) Dancing on DNA: kinetic aspects of search processes on DNA. *Chem. Phys. Chem.* **12**, 1481-1489

21. Vuzman, D., Azia, A. & Levy, Y. (2010) Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails *J. Mol. Biol.* **396**, 674-684

22. Zhou, H. X. (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8651-8656

23. Zinoviev, V. V. et al. (2007) Differential methylation kinetics of individual target site strands by T4Dam DNA methyltransferase *Biol. Chem.* **388**, 1199-1207

24. Porecha, R. H. & Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10791-10796

25. Hedglin, M. & O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. *ACS Chem. Biol.* **5**, 427-436

26. Gowers, D. M., Wilson, G. G. & Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. U. S. A.* **102**,15883-15888

27. Mashhoon, N. et al. (2004) Functional characterization of *Escherichia coli* DNA adenine methyltransferase, a novel target for antibiotics. *J. Biol. Chem.* **279**, 52075-52081

28. Fersht, A. (1998) *Structure and Mechanism in Protein Science*, W. H. Freeman, New York.

29. Coffin, S. R. & Reich, N. O. (2009) *Escherichia coli* DNA Adenine Methyltransferase: the structural basis of processive catalysis and indirect read-out. *J. Biol. Chem.* **284**, 18390-18400.

30. Boye, E., Marinus, M. G. & Lobner-Olesen, A. (1992) Quantitation of Dam methyltransferase in Escherichia coli. *J. Bacteriol.* **174**, 1682-1685

31. Handa, V. & Jeltsch, A. (2005) Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome *J. Mol. Biol.* **348**, 1103-1112

32. Wienholz, B. L. et al. (2010) DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. *PLoS Genet.* **6**, e101106

33. Halford, S. E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* **37**, 343-348

34. Rau, D. C. & Sidorova, N. Y. (2010) Diffusion of the restriction nuclease EcoRI along DNA. *J. Mol. Biol.* **395**, 408-416

VI. Appendix (experimental procedures)

PCR:

Materials:

Plasmid (template): usually about 50ng/µl of a pBR332 derivative (~4300bp)

10X Taq buffer: 500 mM KCl, 100 mM Tris HCl (pH= 8.3), 15 mM $MgCl_2$.

100µM Primers, 25µM DNTPs (each), Taq (lab stock), H_20

Directions:

Make into a 200µL solution (in order from below):

Add water, Taq buffer, 3µL plasmid, 2 µL primers, 4 µL DMSO, then 1.2 µL dNTPs and 7 µL Taq. Vortex then split into 2 PCRtubes.

Notes: thaw (don't vortex) DNTPs and Taq as they are delicate.

PCR program:

1. 95 deg, 1 min
2. 95 deg, 30 sec
3. 60 deg, 30 sec
4. 72 deg, 1 min
5. repeat to 2 for 31 times
6. 72 deg, 2 mins
7. 10 degrees forever

Note: annealing (step 3) depends on the primers, and can be modified (using a gradient program).

Use PCR clean up kit (Agilent, Qiagen)

Ethanol Precipitation:

Add 2.5X of -20 degree 100% ethanol per volume of PCR product. Place in -80 freezer for 20mins or longer. Spin down in table top centrifuge for 30mins at 14,000 RPM. Remove ~90% of liquid (you should see a small pellet). Add ~300 μL -20 degree 70% ethanol, vortex, and then spin at 14,000 for 5 minutes. Remove all liquid by pipette and dry in oven (65 degree) for about 20 minutes. Finally, re-suspend in TE or water.

Processivity assay:

Thaw/ vortex MRB, BSA, DNA, then add to water. Let SAM thaw (don't vortex) and incubate at 37 degrees for ~20 minutes prior to reaction initiation. Typically, I used a 8 μL reaction volume. Incubate 8 μL of water in a 0.6 mL tube in a heat bath at ~75 degrees (for heat kill). Remove Dam from -80, dilute in protein dilution buffer, and initiate the reaction within less than ten minutes from removing from -80. It is ok to vortex dam enzyme before and following reaction initiation. Place 1 μL of reaction aliquot in the 8 μL of 75 degree water for heat kill during reaction.

Following the reaction, turn off heat block following last time point (don't let samples be at 75 degrees for more than ~45 minutes). Let slow cool until ~50 degrees. Remove reactions and add 2 μL DpnII reaction buffer (NEB) and 2 μL of 1/30 DpnII enzyme (dilute in storage buffer (NEB)). Vortex, then place in 37 degree room for > 4 hours but ideally over-night. Cutting tends to be the most problematic step; therefore, introduce controls of unmethylated and completely methylated (reaction at "infinity" time point) DNA.

Following PAGE then densitometry, the values of singly and doubly methylated DNA at each time point are known. They can be placed in the following excel sheet (scheme 1, formulas given). Time is placed in columns B and L. Singly methylated DNA is in column N, doubly methylated DNA is in column M. The solver function is used, where a cell (any,

xx) is the sum of both columns Q and R. Using solver, you "minimize" the solver cell (xx) by changing I4 and J4 (which are k_1 and k_2 respectively). Upon "solving," the appropriate values of k_1 and k_2 are given. The quality of the fit can be checked by overlaying the data on a separate excel sheet where the equations for singly and doubly methylated DNA (see Chapter 1) can be reproduced with the k_1 and k_2 values from the solver fit. Note: the solver driver my need to be downloaded to your computer.

Scheme 1: Excel formula sheet for solver function to get k_1 and k_2 values. Columns A thorough F are on the top section, and columns G through R are below. Here, the spreadsheet only supports an experiment with two time points; this can be expanded for experiments with multiple time points.

	column B	C	D	E	F
row	time			single methyls	double methyls
4	1	=EXP(-I4*B4)	=EXP(-J4*B4)	=((I4)/(I4-J4))*(D4-C4)	=1+(1/(I4-J4))*(J4*EXP(-I4*B4)-(I4*EXP(-J4*B4)))
5	2	=EXP(-I4*B5)	=EXP(-J4*B5)	=((I4)/(I4-J4))*(D5-C5)	=1+(1/(I4-J4))*(J4*EXP(-I4*B5)-(I4*EXP(-J4*B5)))

	H	I	J	K	L	M	N	O	P	Q	R
(row)		k1	k2		time	dm	sm	dm mod	sm mod	dm least	sm least
4					1			=F4	=E4	=(O4-M4)^2	=(N4-P4)^2
5	process	=J4/(I4+J4)			2			=F5	=E5	=(O5-M5)^2	=(N5-P5)^2